THIS WEEK

EDITORIALS

HUE SAYS? Colour firmly in the confused eye of the beholder **p.6**

WORLD VIEW China seeks global leadership in GM food **p.7**

PLANTS Gene disrupter keeps the potato pests away **p.9**

Gone fishing

An investigation into the funding sources of climate scientists who have testified to the US Congress makes demands that have the potential to infringe on academic freedom.

Perhaps it was to be expected. Just days after documents surfaced that raised conflict-of-interest questions about the funding sources of noted climate sceptic Willie Soon, a member in the US House of Representatives entered the fray. On 24 February, Raúl Grijalva, the leading Democrat on the House Committee on Natural Resources, released letters that he had sent to seven universities demanding information on the funding sources of seven other scientists whose views he does not appreciate. Grijalva was right when he wrote in the letters that conflicts of interest "should be clear to stakeholders", but his investigation sends all the wrong messages.

Somewhere behind Grijalva's motives there is a legitimate point. Scientists have a responsibility to disclose their funding sources and any other ties that could be perceived as conflicts of interest when they publish their work. Institutions, including the Harvard-Smithsonian Center for Astrophysics (CfA) in Cambridge, Massachusetts, where Soon works, must establish policies that lay out the rules for their researchers. Scientific journals must also ask authors to declare possible conflicts. These disclosures should apply to funding from industry and from foundations, regardless of which way they lean, as well as from environmental groups. Where there is evidence that these standards are not being met, there is certainly scope to investigate why.

As a result of documents obtained through a US Freedom of Information Act and released last month by environmentalists, the CfA is now reviewing Soon's case and its own policies (see *Nature* http://doi.org/2jx; 2015). This is as it should be, but Grijalva's inquiry is a fishing expedition that seems to have been crafted for publicity rather than clarity. Among his targets are a few long-time climate sceptics, such as Richard Lindzen at the Massachusetts Institute of Technology in Cambridge. Also on the list are policy researcher Roger Pielke Jr at the University of Colorado Boulder, whose 'sin' has been to question political convention on climate issues, and Judith Curry, a climate scientist at the Georgia Institute of Technology in Atlanta who has engaged with climate sceptics.

All of the researchers have testified before Congress, and Grijalva says that his goal is to maintain public confidence in public institutions by ensuring that public policies are not improperly influenced by outside money. Unfortunately, he laments, congressional disclosure requirements did not compel researchers to report their sources of funding, and "we need to fill in those gaps". His letters are addressed to the presidents of the researchers' universities and request information about financial disclosure policies, sources of external funding and any formal disclosures of such funding. They also ask for all drafts of public testimony that the researchers "helped prepare for others" and any communications about the preparation of testimony.

Not only does this investigation shine a high-profile light on researchers before the evidence to judge them has even been gathered, but it goes well beyond questions about funding and disclosure by seeking early testimony drafts and personal correspondence. (Grijalva admitted earlier this week that this was an "overreach", although he is letting his requests stand for now.) A spokesman for Grijalva and the committee's Democratic minority sought to distinguish between this investigation and a 2005 episode in which former chairman of the House Energy & Commerce Committee Joe Barton (Republican, Texas) requested personal communications and scientific data on palaeoclimate research from scientists including Michael Mann, now at Pennsylvania State University in University Park. Grijalva is not seeking scientific data, but there is a reason for the comparison. In both cases, politicians are singling out researchers with whom they disagree and are seeking access to private deliberations that should be protected in the name of academic freedom.

"Politicians are singling out researchers with whom they disagree."

Scientists must view their funding sources as public information that is always subject to scrutiny, and act accordingly. But when politicians seek to probe beyond possible sources of external influence on published work and attempt to expose internal discussions that they find inconvenient, that sends a chilling message to all academics and to the wider public. ■

Fatal fallout

The Ebola epidemic has had a dire effect on the health prospects of pregnant women.

The late stages of pregnancy are a difficult time for most women, but try to imagine what it must be like right now for would-be new mothers in rural areas of Sierra Leone, Guinea or Liberia. Their eight or nine months of pregnancy have already been overshadowed by the ravages of the Ebola outbreak. Now, when they start to feel abdominal cramps, they are faced with an impossible choice.

Before the epidemic, health educators urged pregnant women with complications to report to clinics. But the nearest clinic is typically a journey of a day or more away — and stories abound of friends and relatives who went to the hospital, only to be told that they had Ebola and never come home.

Pregnant women who do brave the journey are often denied care. Some end up delivering their babies alone on floors or in the backs of ambulances. What would you do — would you make the journey?

Now put yourself in the place of the hospital nurse greeting a heavily pregnant woman who arrives at a triage department, weary from her journey, and complaining of abdominal pain. Such pain is, after all, a classic symptom of Ebola, and although the numbers of cases are easing, you have seen colleagues and friends help pregnant women with Ebola,

only to contract the virus and die. The woman's pain worsens and she goes into labour. Would you help? Or, to protect yourself, would you leave her squirming on the floor to deliver the baby herself?

As we explore in a News Feature on page 24, women and health workers in West Africa are facing decisions like these every day. It helps to explain how, as new infections of Ebola are finally being brought under control and the world's attention moves on, the outbreak's devastating impact on maternal health will linger for years.

Pregnant women are uniquely vulnerable to the effects of Ebola, and it is extremely difficult to distinguish the disease's symptoms from routine pregnancy complications. And those who care for these women take their lives in their hands: pregnancy and childbirth necessarily expose carers to potentially infectious bodily fluids. As a result, many doctors, nurses and clinics have refused to treat any pregnant woman who presents with symptoms that could mark her as having Ebola. The United Nations Population Fund (UNFPA) has estimated that the maternal mortality rate — the annual number of maternal deaths per 100,000 live births — may double as a result. And this is happening in countries that already had among the worst maternal-health records in the world.

Some health workers have been brave enough to continue caring for pregnant women during the Ebola epidemic. These include doctors with the medical-aid group Médecins Sans Frontières (Doctors Without Borders) who have devised innovative ways to treat these women and have opened clinics specifically to care for these most vulnerable patients.

Samuel Batty and Amadu Jawara, two Sierra Leonean community health workers, have also stepped up. Both were assigned to work at a hospital in Freetown. When many nurses and doctors abandoned their posts, Batty and Jawara did not.

In November, Batty and Jawara assisted a pregnant woman with a fever. They gave her medication and her fever improved. Assuming that she had malaria, Batty examined her using no special Ebola precautions. It was a fatal mistake. Soon after, Batty himself died of the disease. Even after seeing his friend die, Jawara has continued to care for patients, knowing that they have nowhere else to turn. He estimates that he has performed 100 Caesarean sections.

Community health workers would not usually perform such a procedure. Both Batty and Jawara were trained in surgical skills through a programme run by the Norwegian non-governmental organization CapaCare in conjunction with the Sierra Leonean health ministry. The programme exemplifies an approach called surgical task-shifting, which attempts to redress the dearth of medical personnel in countries such as Sierra Leone by training health workers to give lifesaving care that might otherwise be unavailable to patients in under-resourced areas. CapaCare estimates that, by August 2014, approximately half the surgical procedures performed by its students were emergency obstetric procedures.

"Pregnant women are uniquely vulnerable to the effects of Ebola."

There is debate over the ethics of task-shifting: some worry that it risks exposing patients to substandard care. But the epidemic shows that the people trained by CapaCare are extraordinarily committed to their patients. In this setting, many people faced a choice of being cared for by Jawara, Batty and their counterparts or receiving no care at all.

Sierra Leone had only seven obstetricians before the outbreak began. The UNFPA is seeking US$56 million to fund a new initiative to reopen health services for pregnant women and to recruit more than 500 midwives, doctors and health workers across the outbreak region. International donors should support this initiative.

Training more doctors and finding ways to incentivize them to stay in West Africa are priorities to help the region replace health workers who have lost their lives fighting the epidemic, and task-shifting has proved that it can be part of the solution. The CapaCare programme is currently on hold as a result of the deaths of Batty and another trainee. Restarting medical programmes, including ones such as this, are crucial as the region fights to end the outbreak and begin the long recovery process. ■

Hues and cry

A blue dress divided the Internet — and put the science of visual perception in the spotlight.

The influence on science and the arts of Johann Wolfgang von Goethe — poet, playwright, novelist, proto-scientist, philosopher and general all-round egghead — is profound. His views about the physiological nature of colours, however, have never really caught on, in part because he proposed that colours are more an invention of the mind than a physical reality. One thing, however, rings true: the appearance of objects is not objective, but a conversation between the observer and observed.

Neuroscientists have long recognized that the perception of colour and shade depends strongly on context. Illusions exist, for example, in which one can be utterly convinced that black is white, depending on the surrounding patterns or the conditions in which an object is lit. But it is also true that all other things being equal, the perception of colour differs between people.

One editor of this journal, for example, once owned a car that was, in his opinion, quite clearly green. It remained green in all conceivable circumstances of context, shade and illumination other than complete darkness. Except, however, that everyone else was equally convinced that it was blue — including the vehicle-licensing authority. The car was not only blue — it was officially blue.

Last week, the Internet was deluged with strongly held opinion about colour, specifically of a dress. The dress was advertised as being blue and black. But if illuminated in a certain way, the dress appeared white and gold. People were absolutely convinced of its colour combination, one way or the other. The web exploded with chromatic debate after various celebrities bruited their opinions on Twitter. A straw poll of *Nature*'s editors (including the owner of the blue car) was roughly split down the middle, and convictions were strong — one way or the other.

The explanation for the illusion lies in the colour of the light in which the dress was photographed. The brains of people who read the overall ambience as too blue will overcompensate, seeing the dress as white and gold. Others, whose visual systems read that the lighting was not blue enough, saw the dress as blue and black.

Wired magazine hosted a full discussion on the effect (see go.nature.com/uqf7bo), and the consternation in that publication's office seemed to reflect the brouhaha that briefly reigned in the otherwise serene halls of *Nature*. (The wheels of this international weekly journal of science briefly ground to a halt as so: "I can't read any more manuscripts until I find out WHY?!")

On being told of the illusion, some people — but not all — could just about force themselves to see the dress as black and blue rather than white and gold. The picture is a clear demonstration that colour perception varies between individuals, and according to the conditions of illumination. Such perception is distinct from the genetic conditions that predispose people to the various syndromes known as colour blindness.

Had the ghost of Goethe been watching 'dressgate', he might have allowed himself a rueful smile, given the brickbats thrown in his direction by his scientific critics even in his own time, who, he said, "forgot that science arose from poetry, and did not see that when times change the two can meet again on a higher level as friends." ■

➲ NATURE.COM
To comment online, click on Editorials at: go.nature.com/xhunqv

China's scientists must engage the public on GM

The country's shifting stance on genetic modification for crops needs the support of researchers to persuade a sceptical public, says **Qiang Wang**.

China is about to get serious on the use of genetic modification (GM). After years of uncertainty, funding cuts and public arguments, the country's central government has issued a clear edict: China needs GM, and it will work to become a world leader in the development and application of the technology.

The intent is signalled by the government's first policy document of the year. Issued on 1 February, the state's No. 1 Central Document pledges more government support for research on GM techniques, especially for crops. China has expressed similar enthusiasm for GM technology before, and then backed off in the face of public protest. The policy document shows that the Chinese government does not want that to happen again. The document highlights the need for comprehensive studies to make sure that the technology is safe to use, and it also stresses that Chinese scientists must do more to convince a sceptical public of its benefits.

The new responsibility placed on Chinese researchers to communicate with the public is a significant and positive step forward. It could help to counter the widespread and irrational fear in China that GM food is unsafe to eat. A 2010 online poll of nearly 50,000 Internet users conducted by the news portal *China Daily* found that 84% would not choose GM food for safety reasons. Irrational opinions have sometimes been so strong that scientists have been intimidated and shied away from speaking out. They fear the 'soft violence of violent language' that is too often directed at researchers who simply advocate the commercialization of GM technology.

GM TECHNOLOGY HAS THE **POTENTIAL** TO PRODUCE **MORE FOOD** WITH **LESS POLLUTION.**

China cannot afford to turn its back on GM. Policy-makers are right to include the technology in a new agricultural model for the country. To be self-sufficient, China must grow food for nearly one-fifth of the world's population, with access to just 6% of the world's fresh water and 7% of the world's arable land. In recent decades, China has performed heroics and produced more and more grain — nearly doubling production between 1978 and 2013. But at what cost? The increase was driven by a sixfold rise in the use of chemical fertilizers, which pollute the land and water. Despite China's reputation as a global factory, national reports identify agriculture, not industry, as the biggest source of pollution.

GM technology has the potential to produce more food with less pollution. Without it, China seems likely to become more dependent on imports to feed its people. China imported a record 90 million tonnes of grain last year; cereal imports were up one-third on 2013.

Chinese policy-makers recognize the appeal of GM technology, but must address hostile public opinion. Official announcements on the subject in 2014 even neglected to use the term, preferring instead the euphemism 'molecular breeding'. This might seem surprising given the Chinese government's intolerance of dissent on some issues. But GM technology is seen as a less politically controversial topic, and one with more flexibility for public opinion to steer policy. As a result, China's GM-crop-planting areas have declined since the late 1990s. Brazil and India now grow more GM crops than China. Funding has also been cut: money for major GM-seed cultivation programmes was cut to 400 million yuan (US$63 million) in 2013, down from 2 billion yuan in 2010.

➲ NATURE.COM
Discuss this article online at:
go.nature.com/o1zqn2

Although China imports substantial amounts of GM products — maize (corn) and soya bean among them — it grows only cotton and papaya commercially. Cultivation of GM staple food crops, including rice, is banned, despite the government declaring in 2009 that two GM rice varieties were safe to grow.

The new central document does not introduce specific policies or commitments to boost GM cultivation. These are likely to come later, from specific government agencies. But the decree is an important and influential signal of the state's intentions. Similar documents issued in the early 1980s, for example, accelerated land reforms, and catalysed the dismantling of Mao Zedong's collective ownership of farms. This is widely seen as the start of broader changes in Chinese society, from a planned to a market-based economy.

China needs a similar shift in its attitudes to GM. Despite their enthusiasm for GM technology, and their awareness of its importance for domestic agriculture and food production, the Chinese authorities have so far failed to weigh in on the very public debate over its safety. But it now looks as though that will change.

Most scientists in China work directly or indirectly for the government. Traditionally, these researchers have not been given, or been encouraged to take on, any broader social duties or responsibilities. The latest step could see Chinese scientists talking to the public about GM, as part of a wider effort to overturn scepticism and hostility.

The popularity of social-media sites such as WeChat, which has around 400 million registered users in China, and Weibo (the Chinese version of Twitter) gives researchers a new way to enter into genuine and respectful dialogue over the true risks and benefits of GM. If China is to make good on its intentions to boost its GM efforts, then more scientists speaking up is a good place to start. ■

Qiang Wang *is conjoint professor at the Xinjiang Institute of Ecology and Geography of the Chinese Academy of Sciences in Urumqi, China.*
e-mail: qiangwang7@outlook.com

RESEARCH HIGHLIGHTS

Selections from the scientific literature

PALAEONTOLOGY

Caimans ruled ancient wetland

A vast diversity of specialized crocodilian species dominated the mega-wetlands of South America before the Amazon River flowed, a remarkable fossil find has revealed.

Rodolfo Salas-Gismondi at the University of Montpellier in France and his colleagues found two bone beds in Peru containing seven species of crocodilian — the largest diversity of such species ever found in one place.

As well as two known large-bodied species, they found five animals that are new to science, including several caimans with teeth that seem to be specialized for consuming shellfish. The numbers of these animals declined as Amazon River systems began forming around 10.5 million years ago, draining the wetlands and allowing more-generalist caiman predators to dominate.

Proc. R. Soc. B 282, **20142490 (2015)**

IMMUNOLOGY

Invading bacteria trigger DNA alarm

Immune systems use a previously unrecognized DNA detector to identify invading bacteria.

White blood cells called neutrophils recognize bacterial DNA, triggering a response that eventually kills the invaders. Zusen Fan and his colleagues at the Chinese Academy of Sciences Institute of Biophysics in Beijing found that a DNA-binding protein called Sox2 is also part of this bacterial surveillance system in mice and humans.

They discovered that Sox2 binds to bacterial DNA, and that bacterial infections were worse in mice that had been engineered to have no Sox2 expression in neutrophils. Infections were also worse in mice lacking another protein called TAB2, which interacts with Sox2. The findings could suggest new ways of treating infections, say the authors.

Nature Immunol. **http://dx.doi.org/10.1038/ni.3117 (2015)**

PAUL DOMENICK/GETTY

ZOOLOGY

Eyelash length explained

The optimal length for mammalian eyelashes is one-third of the eye's width, which helps to retain moisture and keep out dust.

David Hu and his colleagues at the Georgia Institute of Technology in Atlanta measured the eyelash lengths of 22 species of mammals and found this ratio. They tested mock-ups of a mammalian eye in a wind tunnel to see how airflow changed depending on eyelash length. Lashes that were shorter than one-third of eye width were not optimal at blocking air from blowing onto the ocular surface. However, longer lashes directed more airflow towards the eye, making it susceptible to drying out. Understanding how eyelashes function could lead to devices that protect optical sensors, the authors say.

J. R. Soc. Interface **http://doi.org/2gk (2015)**

ANIMAL BEHAVIOUR

Birds allow kin to borrow nests

Female ducks recognize their kin and allow them to add eggs to their nests, but fight such attempts by non-relatives.

Many birds try to trick others of the same species into incubating their eggs to avoid the associated energy costs. Malte Andersson at the University of Gothenburg in Sweden and his colleagues studied this 'brood parasitism' by filming the nests of High Arctic common eiders (*Somateria mollissima*) for more than 4,100 hours.

They also analysed the proteins in egg albumen to determine the relatedness of the females that laid eggs in the nest, and found evidence for discrimination against non-relatives.

In 65 nests studied, 11 contained eggs from two different females. At eight of these nests there were fights, and the two females laying eggs in each nest were unrelated. At three nests no aggression was observed, and the laying females were significantly more closely related than in the other eight.

Behav. Ecol. **http://doi.org/2gj (2015)**

BACTERIOLOGY

Altruistic bacteria share their food

Starving bacteria can hook onto other bacterial species to share their nutrients.

Marie-Thérèse Giudici-Orticoni of Aix-Marseille University, France, and her colleagues cultured *Clostridium acetobutylicum*, which uses glucose to grow, and *Desulfovibrio vulgaris*, which uses lactate and sulfate, in a medium containing only glucose. *Desulfovibrio vulgaris* attached itself to *C. acetobutylicum*, allowing it to share the other bacterium's cytoplasm and proteins. This altered the metabolism of *D. vulgaris*, allowing it to grow with only glucose.

In a separate study, Christian Kost of the Max Planck Institute for

Chemical Ecology in Jena, Germany, and his colleagues mutated *Escherichia coli* and *Acinetobacter baylyi* so that they could not produce certain essential amino acids. When grown in a medium lacking the amino acid it required, *E. coli* formed nanotubes up to 14 micrometres long to connect with and share the cytoplasm of nearby *A. baylyi,* which was producing the amino acid. In return, *E. coli* provided *A. baylyi* with the amino acid it needed. These bacteria function as interconnected entities rather than individuals, the authors suggest.

Nature Commun. **6, 6283; 6238 (2015)**

NEUROSCIENCE

Monkeys predict cooperation

Monkeys use a distinct set of neurons to predict whether a fellow primate is likely to cooperate for a common good.

Keren Haroush and Ziv Williams at Harvard Medical School in Boston, Massachusetts, inserted electrodes into the brains of four rhesus macaques (*Macaca mulatta*) to record activity in hundreds of individual neurons in a specific area of the frontal lobe.

They then trained different pairs of macaques to play a computer game displayed on a shared screen. Both animals were rewarded with juice if they cooperated by selecting orange shapes instead of blue ones. A large subset of the animals' neurons fired in a pattern that accurately predicted their partners' intended, as-yet unknown, selections. This subset was distinct from other subsets that fired in patterns reflecting their own personal selections or the expected reward.

Changes in these newly identified 'other-predictive' neurons may be relevant in social behavioural disorders such as autism, the authors say.

Cell **http://doi.org/2gn (2015)**

STEM CELLS

Stem-cell hope for Parkinson's

Dopamine neurons derived from stem cells and inserted into a monkey's brain reduce Parkinson's-like symptoms over two years.

Ole Isacson at Harvard Medical School in Boston, Massachusetts, and his colleagues created dopamine-producing midbrain neurons using induced pluripotent stem cells derived from the skin of cynomolgus monkeys (*Macaca fascicularis*). These neurons were introduced into the brains of three monkeys that lacked dopamine neurons — a model for Parkinson's disease, in which monkeys have impaired motor skills and are less active than normal monkeys.

In the most successfully treated animal, the stem-cell-generated neurons survived and grew axons, and dopamine production was restored. This animal gradually improved during the two years after treatment and showed normal activity, suggesting that transplantation of stem-cell-derived neurons could one day treat Parkinson's disease.

Cell Stem Cell **http://doi.org/2gm (2015)**

AGRICULTURE

Beetles felled by potato RNA

Plants can be engineered to contain molecules that disrupt insect genes, fending off a superpest that is resistant to all major insecticides.

Ralph Bock of the Max Planck Institute for Plant Physiology in Potsdam, Germany, and his colleagues engineered tobacco and potato plants so that their chloroplasts (the cell's photosynthetic structures) expressed RNA molecules that target vital insect genes. Larvae of the superpest Colorado potato beetle (*Leptinotarsa decemlineata*; **pictured**) died after nibbling leaves from the transgenic potatoes. By contrast, potatoes expressing the RNAs outside chloroplasts were not protected — probably because the plant's internal defence mechanism stopped the RNAs from accumulating to sufficient levels.

Science **347, 991–994 (2015)**

MPI CHEM. ECOL.; MPI MOL. PLANT PHYSIOL.

SOCIAL SELECTION *Popular articles on social media*

Psychology journal bans *P* values

A controversial statistical test has met its end, at least in one journal. Earlier this month, the editors of *Basic and Applied Social Psychology* (*BASP*) announced that the journal would no longer publish papers containing *P* values, because the values were too often used to support lower-quality research.

Authors are still free to submit papers to *BASP* with *P* values and other statistical measures that form part of 'null hypothesis significance testing' (NHST), but the numbers will be removed before publication. "Basic and Applied Social Psychology just went science rogue and banned NHST from their journal. Awesome," tweeted Nerisa Dozo, a PhD student in psychology at the University of Queensland in Brisbane, Australia. But Jan de Ruiter, a cognitive scientist at Bielefeld University in Germany, tweeted: "NHST is really problematic", adding that banning all inferential statistics is "throwing away the baby with the p-value".

Basic Appl. Soc. Psych. **37, 1–2 (2015)**

Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➲ NATURE.COM
For more on popular papers:
go.nature.com/ynfi49

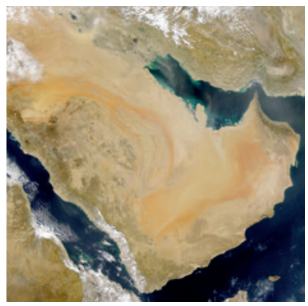
PALAEOCLIMATOLOGY

A damp dispersal out of Africa

Early humans would have had several rainy opportunities to move out of Africa and into the normally arid and challenging Arabian peninsula (**pictured**).

Ash Parton of the University of Oxford, UK, and his colleagues discovered layers of sediments laid down by ancient rivers in southeast Arabia, which flowed for several long periods during the past 160,000 years. Those wet spells could have enabled humans to push into the Arabian interior much earlier than some theories have suggested. Since at least 160,000 years ago, monsoon rains would have provided enough fresh water and plants to sustain human migration approximately every 23,000 years.

Geology **http://doi.org/2f8 (2015)**



SEAWIFS PROJECT, NASA/GSFC; ORBIMAGE

➲ NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

POLICY

Russian ISS plan

Russia's space agency Roscosmos announced on 24 February that it will continue its involvement in the International Space Station (ISS) until 2024 — a timeline that the United States had committed to last year. Roscosmos also added that after 2024 it will consider taking the Russian-built ISS modules and assembling them into a separate space station. Last year, as US–Russian tensions rose over the crisis in Ukraine, Russia's deputy prime minister Dmitry Rogozin said that the country would pull out of the ISS by 2020.

UK embryo law

Mitochondrial donation will become legal in the United Kingdom after the final vote in a debate on 24 February that may set an international precedent. The House of Lords voted overwhelmingly to approve regulations on human fertilization that would allow the creation of embryos with DNA from three people. The technique aims to prevent disease passing from mother to child through the mitochondria — the cell's energy-producing structures, which have their own genes. Only 48 members of the Lords voted against the regulations, and 280 voted for them. The previous vote in the House of Commons was more closely contested (see go.nature.com/hyirxf).

Green-card spouses

Spouses of highly skilled foreign workers will soon be allowed to work legally in the United States, the US government announced on 24 February. The measure, due to take effect on 26 May, will apply to those married to individuals who are in the process of obtaining permanent residency (or a 'green card') while on an H-1B visa — many of whom are scientists or engineers. The US Citizenship and Immigration Services estimate that nearly 180,000 couples will benefit from the policy change in its first year and 55,000 per year after that.

WILL BURRARD-LUCAS/NATUREPL.COM

Chinese pandas increasingly isolated

China's wild pandas have seen an increase in living space, and there are now 1,864 of them — compared with 1,596 a decade ago — all of which sounds like good news. But the results of a four-year survey announced on 28 February are not necessarily cause for celebration, and some experts are still concerned. Although living area has grown, panda populations are increasingly isolated, their habitats fragmented by roads, railways, dams and mines. Climate change threatens their food source, bamboo. And it is not clear that numbers from the latest survey can be directly compared to the previous search around ten years ago. See go.nature.com/h93hle for more.

Pipeline veto

US President Barack Obama vetoed legislation on 24 February that would have authorized the construction of a controversial pipeline intended to carry oil from Canada's tar sands in Alberta to the US Gulf Coast. Republican majorities in both houses of Congress passed the legislation earlier this year, arguing that the Keystone XL pipeline would boost economic development; environmentalists argue that it would increase greenhouse-gas emissions because it promotes a dirty source of energy. The fate of the project now rests with the White House pending an environmental review by the US Department of State, which is expected in the coming weeks.

PEOPLE

Tributes to Spock

Scientists on Earth and in space have paid tribute to Leonard Nimoy, who died on 27 February aged 83. NASA administrator Charles Bolden said that the actor, who played *Star Trek*'s half-Vulcan science officer Mr Spock, was "an inspiration to multiple generations of engineers, scientists, astronauts, and other space explorers", adding

that "he made science and technology important to the story". Astronaut Terry Virts aboard the International Space Station tweeted an image of a Vulcan hand salute from orbit. US President Barack Obama hailed Nimoy as a "supporter of the sciences", saying: "Long before being nerdy was cool, there was Leonard Nimoy."

Guilty plea

A US scientist has pleaded guilty to fraudulent HIV vaccine research. On 25 February, the Southern District of Iowa attorney's office said that Dong Pyou Han, formerly a researcher at Iowa State University in Des Moines, pleaded guilty to two counts of making false statements to the US National Institutes of Health. Han admitted adding human HIV antibodies to rabbit blood to make it appear as if his vaccine had induced an immune response in the rabbits. He will be sentenced in May and faces a maximum of 10 years in prison and a US$500,000 fine.

Cosmic outreach

Astrophysicist Neil deGrasse Tyson (**pictured**), director of the Hayden Planetarium in New York City, received the 2015 Public Welfare Medal from the US National Academy of Sciences, the


THE NEW YORK TIMES/REDUX/EYEVINE

academy announced on 26 February. The medal recognizes Tyson's efforts to excite the public about science. Most notably, he hosted last year's 13-part television series *Cosmos: A Spacetime Odyssey*, inspired by Carl Sagan's classic 1980 *Cosmos* series.

RESEARCH

Ebola trial result

Scientists reported the first positive results from a human clinical trial of a drug to treat Ebola on 25 February. A team led by researchers at the French Institute of Health and Medical Research announced that an antiviral drug, favipiravir, halved the mortality rate among people with low amounts of the Ebola virus in their blood. The death rate in the 40-person trial group was 15%, compared with 30% in the historical control group. But the trial leaders caution that the study numbers are small, among other caveats. The results were announced at the annual Conference on Retroviruses and Opportunistic Infections in Seattle, Washington.

Minnesota review

An external review commissioned by the University of Minnesota has found that "substantial change is necessary" in the way in which the university monitors clinical trials. The report, dated 26 February, states that the university's ethical review committees are not sufficiently staffed or trained, and do not adequately consider the needs of vulnerable research subjects, such as children or people with mental illnesses. The report was intended to address faculty concerns that human-research oversight at the university may not be sufficient after a clinical-trial participant with a psychiatric disorder took his own life in 2004.

Placenta project

A mysterious but crucial organ, the placenta, is getting its day in the sun thanks to a US$41.5-million investment by the US National Institutes of Health (NIH). On 26 February, the agency announced that the Human Placenta Project will fund eight or nine research teams to develop tools to monitor the placenta in real time as a proxy for tracking the developing fetus's health. This could include imaging technologies and ways to detect fetal biomarkers in the mother's blood. Much of the programme's budget is redirected from the NIH's \$150-million National Children's Study, which was cancelled in December. See go.nature.com/ohtjm5 for more.

COMING UP

6 MARCH
NASA's Dawn spacecraft will arrive at the dwarf planet Ceres in the main asteroid belt between Mars and Jupiter.
go.nature.com/edumwi

9–11 MARCH
The University of Tokyo has organized the 1st International Workshop on Topological Electronics at the Okinawa Institute of Science and Technology. This new area looks at the interaction between topology, electrical properties and quantum phenomena.
go.nature.com/ykezpt

FUNDING

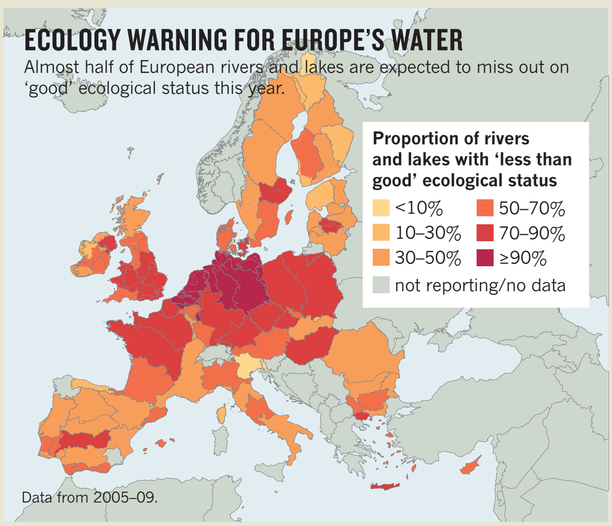
India's budget

Scientists were disappointed by a below-inflation rise for research funding in India's latest budget, announced on 28 February. The overall allocation for science stood at 419 billion rupees (US\$6.8 billion), 3.4% more than was pledged last year. Renewable-energy research was slashed by 68%. But the country's main agency for disbursing research grants, the Ministry of Science and Technology, received an 8% increase, slightly above inflation. See go.nature.com/26qfbq for more.

TREND WATCH

Around half of Europe's rivers and lakes are still polluted, a major environmental review has found, despite a 15-year-old target to restore the continent's waters to good ecological health by 2015. And Germany and the Netherlands are among the worst offenders. The report, by the European Environment Agency in Copenhagen, charts a litany of other environmental failures, including increasing biodiversity loss and continued air and noise pollution. See go.nature.com/cjsoda for more.

SOURCE: EEA



NATURE.COM
For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

ROBOTICS Machines test their mettle in disaster-response contest **p.14**

VOLCANOES Indonesia and the Philippines top list of risky spots **p.16**

CANCER VACCINE A biotech company fails, but its vision lives on **p.17**

A RARE SPECIES Female, a physical scientist and in the developing world **p.20**

MADS NISSEN/PANOS

The trauma caused by death and fear is having long-term ramifications on the people of Sierra Leone.

PSYCHOLOGY

Ebola's mental-health wounds linger in Africa

Health-care workers struggle to help people who have been traumatized by the epidemic.

BY SARA REARDON

The Ebola epidemic in West Africa may be fading, but its impact on mental health could linger for years. Survivors are often haunted by traumatic memories and face rejection by society when they return home, and those who never contracted the disease may grieve for lost relatives or struggle to cope with extreme anxiety.

Aid groups and governments are battling to address the situation in a region that has little in terms of mental-health infrastructure. There has been some progress: on 25 February, for example, the World Bank and the governments of Japan and Liberia announced a US$3-million plan to provide psychosocial support in Liberia. But the fear and distrust of authorities that have helped Ebola to spread also make it difficult to manage the toll on mental health. And measures to contain the virus, such as quarantines, can limit access to the necessary treatment.

"We're still seeing anxiety, and people in survival mode," says Georgina Grundy Campbell, a mental-health nurse with the non-profit International Medical Corps (IMC) in Lunsar, Sierra Leone. "The majority of psychological problems are because the country is frozen, with nothing moving forward," she says.

West Africa is no stranger to crises. In the past two decades, the countries hit by the Ebola epidemic have seen civil war and unrest as well as torture and other human-rights abuses. These events have sparked efforts to improve the nations' limited mental-health-care systems, including a programme funded by ▶

the European Commission in which the first 20 psychiatric nurses graduated from the University of Sierra Leone in 2013. But clinical expertise is still scarce in the Ebola zone: Liberia has just one psychiatrist; Sierra Leone has none.

Even the simplest interactions between people with Ebola, their families and health-care workers are complicated by the precautions needed to prevent infection. Because doctors and nurses can wear their heavy personal protective equipment only for short periods, they focus on providing treatment. Tasks such as counselling bereaved families are often left to mental-health providers from aid groups. In Sierra Leone, for instance, the non-governmental group Community Association for Psychosocial Services (CAPS) has redirected its 18 employees from assisting war survivors to helping people with Ebola and educating communities about the disease. "In this emergency, everyone's kind of in slow motion, making sure that the health staff are safe," says Cynthia Scott, a psychologist with Médecins Sans Frontières (also known as Doctors Without Borders) who recently returned from Sierra Leone.

Surviving the virus presents its own challenges. Some patients refuse to eat or leave their beds. Many blame themselves for contracting the disease. And those who return home are often barred from housing complexes or workplaces. That is a distinct contrast from the way in which communities hit by war or natural disasters typically rally around victims, says Inka Weissbecker, psychosocial adviser at the IMC in Washington DC.

Non-governmental organizations are working to decrease the stigma using approaches such as portraying Ebola survivors as heroes. They are also addressing another contentious issue — regulations that outlaw traditional burial rites — by providing families with photos of their loved ones' bodies, which offer some comfort. "I have heard people say, 'If we cannot bury our people properly, we feel our community is sick,'" Scott says.

"The majority of psychological problems are because the country is frozen, with nothing moving forward."

But some actions taken to limit Ebola's reach are harder to deal with. Fear of spreading infection among doctors and patients prompted the E. S. Grant Mental Health Hospital in Monrovia — Liberia's only such facility — to cease most of its operations last autumn. The facility has discharged most of its patients, including several dozen with psychotic conditions. "There's no doubt there's an increase in the number of people in the streets because the hospital is still not functioning at the normal level," says Benjamin Harris, Liberia's only psychiatrist.

Efforts to build treatment capacity in West Africa are showing encouraging signs. The programme in Liberia will, over a three-year period, deploy mental-health clinicians in schools, among other actions. The Liberian Ministry of Health lists mental health as a priority in its Ebola-recovery plan, along with issues such as maternal care (see page 24) and HIV. And in Sierra Leone, CAPS has treated roughly 1,500 people affected by Ebola.

These developments are part of a broader shift in the global health community's attitudes toward mental health. The World Health Organization increasingly addresses psychological care in its reports, and donors to groups such as the IMC are becoming more amenable to supporting mental-health programmes. But there is still much work to be done to ensure that psychological care is a priority in the Ebola response, Weissbecker stresses. "We have to be vigilant about this," she says, "and make sure it stays on the radar." ■

TECHNOLOGY

Robo-rescuers battle it out

Nimble bots rise to the challenge in DARPA competition.

BY BOER DENG

When the humanoid robot SAFFiR gets a shove, it reflexively moves to maintain its balance. SAFFiR can also walk over uneven terrain, turn its head to scan its surroundings and — with the help of a human operator — reach out to grasp objects.

Built by a team at Virginia Tech in Blacksburg, SAFFiR is a firefighting robot and a prototype for one that will compete in the final stage of the DARPA Robotics Challenge (DRC), a contest run by the US Defense Advanced Research Projects Agency. The aim is to produce robots with improved mobility, autonomy and responsiveness to human commands. On 5 March, DARPA will announce the 25 finalists who will vie for the US$2-million first prize at the final event in June.

Most advanced robots today follow preset instructions in familiar environments, such as a factory floor, or are almost entirely remotely controlled, says Gill Pratt, a programme manager at DARPA who is running the contest. But there is a need for shrewder, nimbler machines that can operate in less-predictable situations.

A robot developed by Japanese company Schaft takes the DARPA Robotics Challenge in 2013.

AP/PA

For example, radiation from the 2011 Fukushima nuclear meltdown in Japan made it unsafe for people to manage the contamination. The robots that could be deployed for clean-up were slow and maladroit, and could do little more than survey the damage.

DARPA set up the contest to spur the development of robots that might perform better in future disasters, and several rounds of the competition have been held since the event's launch in 2012.

A key to winning will lie in a robot's ability to do things on its own, such as deciding how high to lift a leg to climb a step, or tracking its location relative to a target object. During trials in December 2013, robots had to complete tasks such as shutting off a valve, climbing a ladder and driving a car through a winding speedway course. Those with greater autonomy performed better, according to an analysis by Holly Yanco, a computer scientist at the University of Massachusetts at Lowell. Robots were faster and errors were easier to fix when they needed less human input.

For the final round, the machines will compete with a time limit, without an external

BERTIL ENEVAG ERICSON/TT/EPA/CORBIS

power source and sometimes without communication with their operators. This mimics real disaster scenarios, in which connectivity can be spotty. DARPA expects that this constraint will encourage teams to increase their robots' capacity to map out, plan and act independently. "Success at this level would be a huge achievement," says Pratt.

Robots have become reasonably good at sensing and moving, says Henrik Christensen, chair of robotics at the Georgia Institute of Technology in Atlanta. Remote sensors on SAFFiR, for example, measure force and body position (part of what allows the robot to find its centre of gravity); cameras and software let it monitor distance and potential obstacles. CHIMP, a robot built by a team at Carnegie Mellon University in Pittsburgh, Pennsylvania, transitions deftly from trundling along on four tank treads to standing on two.

But the machines are "not much good at making judgements, like deciding if something is a drill or a cup, or figuring out which is relevant to what it needs to do", Christensen says. "These things turn out to be very hard, and humans are much better at them."

DIVISION OF LABOUR

Some teams want to better translate what a robot sees into something that a human operator can easily understand. The goal is to create an efficient human–robot team that is "more like having two humans, where one directs the other", says Todd Danko, who leads a DRC team from defence contractor Lockheed Martin. In fact, Yanco's analysis found that teams with better-designed human–robot interface platforms were better at completing contest tasks.

"I think a big part of this contest is that it will get people to embrace a kind of 'shared autonomy' between the humans and the robots," says James Kuffner, a roboticist at Google in Mountain View, California. "That means thinking about what's necessary for the human to do, for the robot to do, and for how to tell that to the robot."

DARPA-sponsored robot contests in the mid-2000s focused on autonomous transport and were driven by an interest in taking human couriers off the explosive-strewn streets of Iraq and Afghanistan. These races helped to spawn interest in driverless cars from companies such as Google.

In more recent years, the defence department has recommended that the military change its approach to autonomous systems and focus on collaboration between people and technology. Unlike autonomous trucks and unmanned drones, the robots under development for the current competition will be accomplices — not mere substitutes — in helping humans to get difficult jobs done. ■

Former IPCC head Rajendra Pachauri (left) consults with potential successor Thomas Stocker.

POLICY

UN climate panel charts next steps

Intergovernmental Panel on Climate Change prepares for new leadership and another assessment of climate science.

BY JEFF TOLLEFSON

Despite calls for change, the next United Nations climate assessment will take much the same form as the last one, the panel charged with producing the recurring reports announced on 27 February. The decision comes just days after the panel's long-time leader resigned in the middle of a sexual-harassment investigation.

Meeting in Nairobi from 24 to 27 February, the United Nations Intergovernmental Panel on Climate Change (IPCC) made several minor adjustments to its assessment process. The changes aim to engage more scientists, in part by boosting the representation of developing nations in the group's governing body. But the basic framework will continue to comprise a comprehensive assessment published every five to seven years plus two or three special reports on specific topics. The fifth and most recent IPCC climate assessment, which was completed last year, concluded that it is "extremely likely" that humans are responsible for the bulk of recent global warming.

"The overall structure remains, but some key aspects of its mode of operation have been improved to facilitate a fuller participation of all scientists, in particular from developing countries," says IPCC vice-chair Jean-Pascal van Ypersele, a climatologist at the Catholic University of Louvain in Louvain-la-Neuve, Belgium. "This was a key thing I think the IPCC needed to do."

The meeting follows the sudden departure of Rajendra Pachauri, who has headed the IPCC since 2002 and whose term was due to end in October. Pachauri is under investigation over allegations that he sexually harassed a colleague at the Energy and Resources Institute in New Delhi, of which he is director. He has denied the claims but elected to step down on 24 February, soon after announcing that he would not be attending the Nairobi meeting.

"We cannot ignore the resignation of Dr. Pachauri, but the allegations against him … do not relate to the IPCC," said IPCC secretary Renate Christ during a press conference on 27 February. Christ said that the panel will, however, ensure that it maintains an atmosphere in which "everyone's rights are respected and upheld".

Ahead of the meeting, some scientists involved in the IPCC argued that the ▶

assessment process is too slow and requires too much time from the more than 2,000 scientists from around the world who volunteer for duty. Some have advocated that the IPCC put less energy into monumental assessments and more into shorter reports that focus on major scientific and policy debates. During the last major assessment, the IPCC released special reports on renewable energy and the risks of extreme weather, but even those were major undertakings.

Christopher Field, co-chair of the working group on impacts and adaptation for the most recent assessment, says that there are ways to streamline the process, but maintains that the value of the IPCC comes from the give and take between scientists and governments. "Operationally, it is hard to imagine a way to capture this unique value without key process steps, including multiple rounds of monitored review and line-by-line approval of summaries for policy-makers," he says.

OPEN UP

At the meeting, IPCC members said that the next assessment should have a greater focus on specific regions and include a broader review of non-English scientific literature, with more involvement of science writers and communications experts to help reach a broader range of people.

The panel also wanted to open itself up to researchers who have been seeking access to the closed-door meetings in an effort to study the assessment process and the institution itself; research proposals will be evaluated on a case-by-case basis.

"That is indeed a major step forward toward both increased transparency of the IPCC process and eventually finding ways to improve it," says Michael Oppenheimer, a climate scientist at Princeton University in New Jersey who is part of a team of researchers seeking such access.

Oppenheimer has advocated reforms that would emphasize smaller, faster assessments while decreasing the workload for scientists. He says that the latest decision largely represents "business as usual", but does open the door for improvements. In particular, he credited the IPCC for emphasizing communications and engagement with developing countries. "This is important and needs to be done," he says.

The IPCC will hold its leadership election in October. Candidates include van Ypersele and Thomas Stocker, a climate scientist at the University of Bern who co-led the working group that wrote the physical-science portion of the report during the most recent assessment. Field, who is founding director of the department of global ecology at the Carnegie Institution in Stanford, California, says that he, too, is likely to run. ■

Mount Merapi, the most active volcano in Indonesia, erupts every few years.

OSCAR SIAGIAN/REDUX/EYEVINE

NATURAL HAZARDS

Global volcano risk quantified

UN assessment aims to save lives by aiding planning.

BY ALEXANDRA WITZE

Swept away by mudslides, entombed in lava or suffocated under ash, nearly 280,000 people have died in volcanic eruptions during the past four centuries, but only now has humanity managed to quantify the risk posed by these fiery phenomena. The first detailed assessment of global volcanic risk — part of a larger international hazard assessment released on 4 March by the United Nations Office for Disaster Risk Reduction — aims to save lives by providing better information for risk planners and by showcasing effective response measures.

"For the first time, we really have a shared understanding of volcanic activity at the global scale," says Jean-Christophe Komorowski, a volcanologist at the Institute of Earth Physics in Paris, who contributed to the report. "This is a major turning point."

Eight hundred million people live within 100 kilometres of a volcano that could erupt. But the hazards differ greatly from place to place. High in the snow-capped Andes, an eruption might melt ice and send floodwaters rushing into nearby villages. In southeast Asia, a powerful eruption might blast ash over a wide area, causing roofs to collapse under the weight.

The report aims to put hard numbers on exactly who is at risk. It comes from a UK-led international network of institutions called the Global Volcano Model, working with the International Association of Volcanology and Chemistry of the Earth's Interior.

Team leaders sifted through a database of nearly 9,500 eruptions over the past 10,000 years kept by the Smithsonian Institution in Washington DC. They noted how often a particular volcano had erupted and what kind of physical hazards it posed. Then they tallied the number of people who now live within 10, 30 and 100 kilometres of that volcano and whether they live in places where eruptions have killed people before (see 'Mass destruction'). The result is a complete catalogue of the highest-risk volcanoes and a list of countries ranked by the number of residents in harm's way.

Researchers were surprised to find risk in places not typically thought of as highly volcanic. The Auvergne region of France, for instance, has been quiet in historic times. But it has had eruptions in the past few thousand years, putting it relatively high on the hazard scale because so many people live nearby. In New Zealand, the Auckland volcanic field — the eruptive history of which is not particularly well known — lies directly under the country's biggest city.

"Volcanoes are extremely attractive areas to

live," says Jenni Barclay, a volcanologist at the University of East Anglia in Norwich, UK. Volcanic soil is often fertile and the altitude provides good living conditions in hotter climates.

Worldwide, 62 volcanoes fell into the highest risk category, meaning that they have been recently active and lie close to a lot of people. Indonesia tops the list of most threatened countries, with 77 historically active volcanoes, including Mount Merapi, which erupts frequently near the city of Yogyakarta.

But by a different measure, small volcanic islands — such as Montserrat in the Caribbean — are the most vulnerable. When these island nations start to rumble, all their citizens must flee or risk death. In these places, uncertainty has its cost: a controversial 1976 evacuation of the Caribbean island of Guadeloupe left residents angry when no major eruption happened.

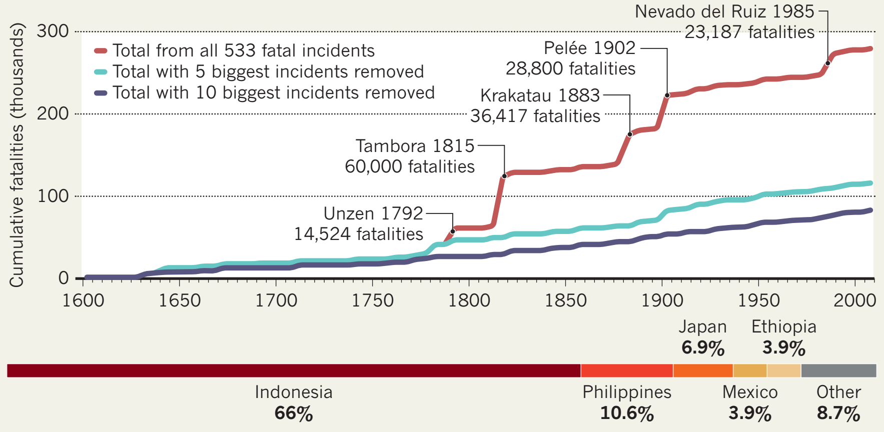
Just because a volcano ranks as hazardous does not mean that people living near it are sitting ducks. If a volcano has enough scientific monitoring equipment on it, and a well-organized local response, then the risk to human life can be reduced, says Stephen Sparks, a volcanologist at the University of Bristol, UK, and a lead author of the report. At Merapi in 2010, the authorities used information about physical changes in the volcano to evacuate hundreds of thousands of people before a large eruption, saving many lives.

"We want to showcase what volcanologists around the world are doing," says Sue Loughlin, a volcanologist at the British Geological Survey in Edinburgh and another leader of the survey.

In Ecuador, around the Tungurahua volcano, local volunteers serve as a network of *vigías* or 'volcano watchers'. They watch for changes in the mountain and radio in to the nearby volcano observatory every evening with their reports (J. Stone *et al. J. Appl. Volcanol.* **3,** 11; 2014). Such initiatives could translate to other volcanically active regions, says Barclay. "We can learn much more by bringing all this knowledge together." ■

MASS DESTRUCTION

More than half of the fatalities caused by volcanic eruptions in the past four centuries occurred in just five major events that killed an estimated 162,928 people. Today, more than 90% of the volcanic risk is concentrated in five countries.



SOURCE: GLOBAL VOLCANIC MODEL/IAVCEI

BIOTECHNOLOGY

Therapeutic cancer vaccine survives biotech bust

Pharmaceutical company rescues landmark prostate-cancer treatment, Provenge.

BY HEIDI LEDFORD

The first therapeutic cancer vaccine to be approved in the United States will stay on the market despite the financial collapse of the trailblazing biotechnology company that developed it. The vaccine, Provenge (sipuleucel-T), was purchased on 23 February by Valeant Pharmaceuticals of Laval, Canada, which paid US$415 million for the prostate-cancer treatment and other assets of the bankrupt Dendreon Corporation.

The now-defunct Dendreon, of Seattle, Washington, made history in 2010 by showing that complex treatments made fresh for each patient could win regulatory approval, and could be expanded beyond the realm of specialized academic hospitals. Industry took note: today, experimental cancer therapies that spur patients' immune cells to attack tumours are among the hottest properties in biotechnology.

"Dendreon had vision and foresight," says Usman Azam, head of cell and gene therapies at Novartis, a Swiss pharmaceutical company that has purchased one of Dendreon's manufacturing plants to fuel its own cell-therapy efforts. "Don't view Dendreon as a failure: it paved the way."

But although Dendreon created the market for such cell therapies, it ultimately could not survive in it.

Provenge is made by harvesting a patient's dendritic cells — a type of immune cell — and then mixing them with a protein that is particularly abundant in prostate tumours. This primes them to recognize and attack the tumour; the cells are then infused back into the patient.

The technique was pioneered in the early 1990s by Edgar Engleman, an immunologist at Stanford University in California, who had seen promising results in animal studies of a different cancer, lymphoma. He teamed up with fellow Stanford immunologist Samuel Strober to work out ways to make the process more efficient.

When the two pitched their idea for a company to investors, they had few clinical data and were too optimistic about how fast the treatment could reach patients, says Strober. The company was an enormous gamble: harnessing the immune system to fight cancer was still a controversial idea, and no other company had marketed a therapy so personalized and labour-intensive. "But at that time it was a little different from now," says Strober. "Companies were getting funded on the basis of promise, rather than actually looking at their capacity for early commercial success."

Engleman and Strober founded Dendreon in 1992; the US Food and Drug Administration approved Provenge in 2010.

The approval was celebrated as an important proof of concept by researchers working on cancer vaccines and other treatments that stimulate immune responses to the disease. But Dendreon, already strained by the long wait for approval, soon ran into financial difficulty. ▶

▶ The United States' publicly funded Medicare system decided in 2011 to pay for Provenge treatment. But confusion over how the cost of the vaccine would be reimbursed by private insurance companies left many US doctors hesitant to use it, says Corey Davis, an analyst at Canaccord Genuity, an investment bank based in Toronto, Canada. When revenues came in far below the company's initial estimates, Dendreon failed to adjust its operations accordingly, Valeant chief Michael Pearson told investors on 23 February.

Provenge is, at first glance, an odd purchase for Valeant, a company known for acquiring relatively simple, established products — for example, it controls 10% of the US contact-lens market. But Valeant saw an opportunity to cut costs and improve how the vaccine is marketed to doctors, and thinks it can make back its investment in less than two years, says Davis.

The vaccine's rescue is a relief to Engleman, who had feared that Provenge might disappear along with Dendreon. As the company struggled financially, the scientists who founded it watched helplessly from the sidelines. "This was our baby," says Engleman. "It was extraordinarily frustrating. There was nothing we could do."

In retrospect, Engleman says, some early scientific choices may have exacerbated Dendreon's struggle. He notes that the company decided not to develop ways to freeze the stimulated immune cells, which could have simplified the procedure and lowered its cost.

"Don't view Dendreon as a failure: it paved the way."

Both scientists lament the choice of prostate cancer as the inaugural disease target of the technology. Although the early lymphoma data had been very promising, recalls Engleman, the company decided to switch to a more common cancer with a bigger potential market. And prostate cancer had another advantage: people can live without a prostate, which helped to calm fears (since proved unfounded) about what would happen if the primed immune cells attacked healthy tissue.

But the results in prostate cancer were not as dazzling as Engleman had hoped on the basis of his animal results in lymphoma. Dendreon did extend survival in some people with advanced prostate cancer, but by a median of only four months (P. W. Kantoff *et al. N. Engl. J. Med.* **363,** 411–422; 2010). Last week, the UK National Institute for Health and Care Excellence advised that at more than £47,000 (US$73,000) per course of treatment, Provenge is too expensive to justify its use by the National Health Service.

The Dendreon experience has not dampened Engleman's enthusiasm for entrepreneurship. He and Strober, along with other collaborators, have teamed up on a company that aims to develop a technique to reduce the likelihood that recipients of transplanted organs will develop an immune response to the new tissue.

They are again on the hunt for funding, but this time the team is backed by more than a decade of clinical-trial data that supports the method. "We're thinking that this one will progress a lot faster than the Dendreon thing," says Strober. ■

HYDROLOGY

Slick idea proposed to stretch water supplies

Thin coating that cuts evaporation from lakes offers hope for drought-ridden United States.

BY MATTHEW WALD

In the southwestern United States, where years of drought are leading water managers to consider drastic water-provision measures such as desalination and cloud seeding, entrepreneurs have suggested reviving a technique that was tried and abandoned half a century ago. They propose to stretch dwindling water supplies by slowing down evaporation from reservoirs by means of a surface barrier of cheap, non-toxic, biodegradable chemicals just one molecule thick — two-millionths of a millimetre. The technology is far from proven, but it showed some potential in field tests in Texas last year.

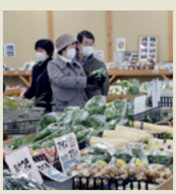
Worldwide, more water evaporates from reservoirs than is consumed, and the losses are especially acute in hot, dry regions. The idea of a coating to slow evaporation has been floated for decades, and government researchers in the United States and Australia have investigated the concept. The approach, which generally involves chemicals derived from coconut or palm oil, is already used on small bodies of water such as golf-course ponds and swimming pools. But it has not been practical for larger bodies because wind tends to break the layer apart, says Moshe Alamaro, an engineer at the Massachusetts Institute of Technology in Cambridge.

Last summer, a field test in Texas attempted to overcome that problem. The US$325,000 test ran from July to October on Lake Arrowhead, a 21-square-kilometre reservoir that serves the city of Wichita Falls. Flexible Solutions International, a company in Victoria, Canada, that makes evaporation-reducing coatings for small water bodies, programmed a boat to run on autopilot and make a grid pattern across Lake Arrowhead spreading the coating behind it.

An analysis published in January by the Texas

MORE ONLINE

TOP NEWS



NATSUKI SAKAI/AFLO/ALAMY

Japanese monitoring effort tracks food radioactivity at Fukushima go.nature.com/lv3rgr

MORE NEWS

- Mapping a giant virus in 3D go.nature.com/xzsp6s
- Ancient DNA reveals how wheat came to prehistoric Britain go.nature.com/dradnn
- Tropical forest losses outpace UN estimates go.nature.com/ozbvuj

Q&A

CONTRASTO/EYEVINE

The science behind the world's first stem-cell therapy go.nature.com/rpbyp3

Drought-stricken Lake Arrowhead in Texas was coated with a film that reduced evaporation by up to 15%.

LARRY SMITH/EPA/CORBIS

Water Development Board suggested that the results were promising, but not conclusive (see go.nature.com/zrcozd). Compared with a similar reservoir nearby, the treated reservoir lost an estimated 15% less water to evaporation. But the analysis could not attribute all of that difference to the coating, because it did not account for variables such as stream inflows and seepage outflows. The coating probably helped, says Mark Wentzel, a hydrologist for the Texas Water Development Board and co-author of the report, but "I wouldn't stake my life on it."

Alamaro suggests that a more aggressive technological approach is needed. Radar instruments carried on a blimp or drone could reveal where the reservoir's coating has broken up, he says, by sensing the way that it dampens ripples on the water. More coating can be added when and where it is needed, potentially cutting evaporation by 70%, he estimates.

Alamaro has founded a company, More Aqua in Cambridge, Massachusetts, that is trying to develop a system of diffusers and skimmers to keep a body of water thoroughly and continuously covered.

The company is planning its own pilot test near Palo Alto, California, this summer using its own coating. It plans to offer its services in exchange for owning the water that it saves, and to sell that on the open market in the state, where the cost of water for irrigation can exceed $1,000 per acre-foot (1,233 cubic metres).

Assuming evaporation savings of 15%, says Daniel O'Brien, president of Flexible Solutions, the $160 cost of saving one acre-foot of water with the coating compares favourably with a Texas market price of $345–700 per acre-foot of water.

Water consultant William Mullican of Lubbock, Texas, who is retired from the Texas water board, says that although the Lake Arrowhead results were unclear, this is also often the case in field tests of other water-sparing techniques such as cloud seeding and cutting back brush to keep it from sucking moisture from the soil.

Given the Texas test's promising result and the fact that the ongoing drought in the state is severe, he says, there is every reason to try the technique out again, "unless it starts raining". ■

CORRECTION

The News Feature 'The painful truth' (*Nature* **518**, 474–476; 2015) omitted Marks' first name and affiliation. Donald Marks is the co-founder and chief science officer of Millennium Magnetic Technologies.


Drought-stricken Lake Arrowhead in Texas was coated with a film that reduced evaporation by up to 15%.

LARRY SMITH/EPA/CORBIS

Water Development Board suggested that the results were promising, but not conclusive (see go.nature.com/zrcozd). Compared with a similar reservoir nearby, the treated reservoir lost an estimated 15% less water to evaporation. But the analysis could not attribute all of that difference to the coating, because it did not account for variables such as stream inflows and seepage outflows. The coating probably helped, says Mark Wentzel, a hydrologist for the Texas Water Development Board and co-author of the report, but "I wouldn't stake my life on it."

Alamaro suggests that a more aggressive technological approach is needed. Radar instruments carried on a blimp or drone could reveal where the reservoir's coating has broken up, he says, by sensing the way that it dampens ripples on the water. More coating can be added when and where it is needed, potentially cutting evaporation by 70%, he estimates.

Alamaro has founded a company, More Aqua in Cambridge, Massachusetts, that is trying to develop a system of diffusers and skimmers to keep a body of water thoroughly and continuously covered.

The company is planning its own pilot test near Palo Alto, California, this summer using its own coating. It plans to offer its services in exchange for owning the water that it saves, and to sell that on the open market in the state, where the cost of water for irrigation can exceed $1,000 per acre-foot (1,233 cubic metres).

Assuming evaporation savings of 15%, says Daniel O'Brien, president of Flexible Solutions, the $160 cost of saving one acre-foot of water with the coating compares favourably with a Texas market price of $345–700 per acre-foot of water.

Water consultant William Mullican of Lubbock, Texas, who is retired from the Texas water board, says that although the Lake Arrowhead results were unclear, this is also often the case in field tests of other water-sparing techniques such as cloud seeding and cutting back brush to keep it from sucking moisture from the soil.

Given the Texas test's promising result and the fact that the ongoing drought in the state is severe, he says, there is every reason to try the technique out again, "unless it starts raining". ■

CORRECTION

The News Feature 'The painful truth' (*Nature* **518**, 474–476; 2015) omitted Marks' first name and affiliation. Donald Marks is the co-founder and chief science officer of Millennium Magnetic Technologies.

THE MINORITY

BY KATIA MOSKVITCH

Patchanita Thamyongkit was waiting patiently near the stage at a conference on the importance of science for Thailand, when the organizer rushed up to her and asked whether she had seen the next presenter, Professor Thamyongkit. "That's me," she replied. An awkward pause followed. "Oh, I thought you were his secretary," came his reply.

The presenter was probably more embarrassed by the 2008 incident than Thamyongkit, who is used to being taken for a secretary. She is a physical organic chemist at Thailand's biggest scientific establishment, Chulalongkorn University in Bangkok, where she has won several awards for her work. But senior female scientists are a rarity in Thailand: top science positions are scarce, and many women are forced out of research because of cultural expectations that they will take care of their households, raise children and help ageing parents.

Women around the world continue to face major challenges in pursuing a research career — particularly in the physical sciences. In the United States, women make up nearly half of college-educated workers and working life scientists, but only 30% of physical scientists. The gender gap widens higher up the professional scale. There is much debate about why: discrimination, conscious and unconscious bias in hiring and promotion, lack of role models and the demands of bearing and caring for children are all thought to play a part.

In developing countries, female scientists can face even higher cultural and societal barriers, such as overt sexism, and a lack of contraception, reproductive choice or access to education. Many live in regions with desperate poverty, few high-quality schools, political instability and sometimes civil conflict. "There is a great, growing interest in science and engineering among women in developing countries," says Romain Murenzi, a physicist and former Rwandan science minister who is now executive director of TWAS, the World Academy of Sciences, which promotes the advancement of science in developing countries. Yet, he says, "women face obstacles from their earliest years, and across the landscape of their lives". The situation varies enormously from one place to the next: figures from the United Nations Educational, Scientific and Cultural Organization (UNESCO) show that women in Thailand make up just over half of science doctoral students and researchers, whereas in Nigeria, those figures are 24% and 23%.

Nature talked to three women who are in this minority of minorities — working in physical sciences in developing countries — to find out what challenges they have faced and how they overcame them. Their stories highlight two of the ingredients for success: a supportive family and a huge dose of determination. Becoming a leading scientist involves "convincing your family first", says Thamyongkit, "and then, for the rest of your life, going out of your way to prove to everyone that you're actually really good".

THAILAND

Patchanita Thamyongkit

THAMYONGKIT LIVES WITH HER ELDERLY PARENTS in a narrow, three-storey house in blaring central Bangkok. The neighbourhood is a sea of concrete, with wall-to-wall houses. When she was young, her parents ran an automobile-parts shop and a book store. They were ambitious for their daughter, and wanted her to become a doctor. "I am lucky in the sense that my parents have always supported me," says Thamyongkit — even when she chose to study physical sciences instead of medicine at Chulalongkorn University.

Thamyongkit embarked on an academic world tour: she did a PhD in Germany and took a postdoctoral post in the United States before returning, speaking two foreign languages, to join the chemistry department at Chulalongkorn in 2005. She researches novel organic materials for optoelectronic applications, such as energy-efficient solar cells. Last November, she won a L'Oréal–UNESCO prize for women in science.

Equal proportions of men and women participate in Thai science education and scientific research as a whole. At Chulalongkorn, roughly 25% of faculty members are women in physics and 35% in chemistry, proportions roughly similar to those in the United States. Many universities have nurseries and schools on campus, and some researchers say that they do not perceive a gender gap at all. The country elected its first female prime minister, Yingluck Shinawatra, in 2011.

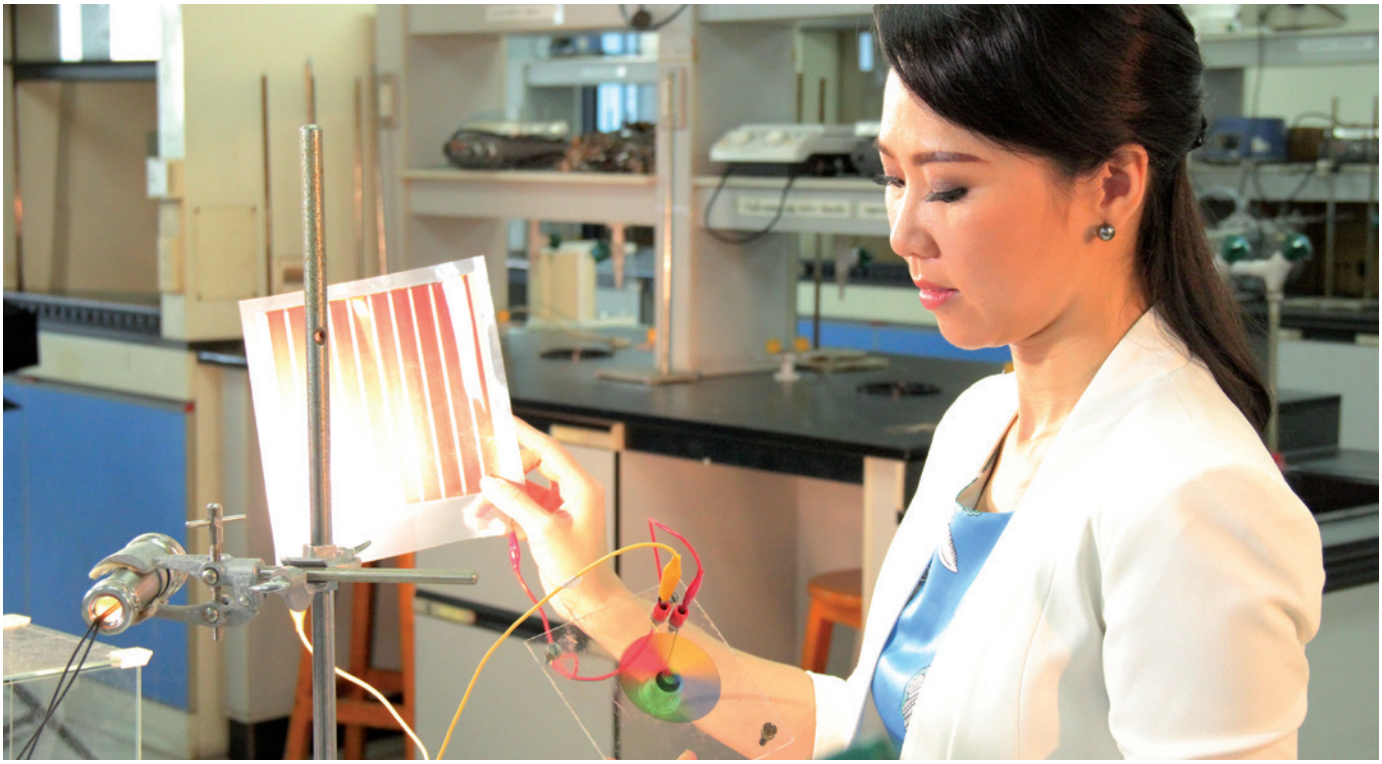
But as in other nations, the academic pipeline for women is leaky. "Really, really few actually reach the top," says Supot Hannongbua, dean of the faculty of science at Chulalongkorn. "Researchers maybe publish one paper a year — it's very difficult for them." Out of 80 faculty members in the department of chemistry, he says, there may be between 5 and 7 people doing outstanding research or in senior positions, and rarely is one a woman.

Part of the problem is that top research jobs are rare. Just 0.2% of the country's gross domestic product goes on research, and only around 20 of Thailand's 160 or so universities have proper research facilities. With so few opportunities, women often lose out. Most of the 12 female graduate students in Thamyongkit's lab late last year said that they wanted to work in industry, care for their families, or both.

That efflux is also driven by Thai society's traditional expectations. Many women work, but it is assumed that they will get married and shoulder caring and domestic duties. Although overt sexism is frowned on, discrimination persists. "If a man and a woman candidate with exactly the same characteristics apply for a top scientific position, the

MINORITY

Women are under-represented in physical sciences and in science in the developing world. Meet three who beat both sets of odds.



Patchanita Thamyongkit is a physical organic chemist at Chulalongkorn University in Bangkok.

man is more likely to get hired," says Hannongbua. Some advertisements for industrial jobs state that the applicant must be a man.

Women can also struggle to make the right connections, says Tyrell Haberkorn, who studies Thai politics and history at the Australian National University in Canberra. "Women in Thailand face everyday sexism of the boy's-club variety, in which important decisions and relationships are forged outside the workplace and instead in the male-only spaces in which men go out and eat and drink together." One female chemist at Chulalongkorn, Parichatr Vanalabhpatana, says that she had to learn to drink beer to be accepted by her male colleagues.

The Thai government has tried to institute change. Shinawatra promoted gender equality through a National Development Fund for Women, which offers low-interest loans to further women's welfare and occupations; and in 2007, Thailand criminalized domestic violence.

But Thamyongkit, Hannongbua and others say that the government should do more. "I still insist that Thailand has no particular and serious campaign to encourage women in science," says Thamyongkit. Overcoming entrenched societal attitudes "would take what would amount to a social revolution in how gendered roles are conceived", says Haberkorn. "But in truth, I think that continuing to open up education, particularly tertiary education, to a broader and broader group of people is the most concrete strategy."

NATURE.COM
For a video of Thamyongkit, see: go.nature.com/zsausd

"Thailand has no particular and serious campaign to encourage women in science."

Thamyongkit has found her own way through. In her twenties, she decided to not have a family. "I did want to have kids before, but if ten years ago I had had a family, I would have never managed to get to the same level in my career where I am now," she says. Now 38, she says that it is probably too late. As she walks through the campus, Thamyongkit passes a temple, where she takes a few seconds to stop, kneel and pray. "I'm not religious," she says. "But I meditate, I say a few words to reinforce all the positive things in my life, to keep the positive energy flowing."

At the end of the day, as Thamyongkit sits down with her family for dinner, her parents say that they would love to have a grandchild. But they don't pressure her, she says, and she is grateful for that. "My relatives want to know why I don't have a family. They say, 'You don't look that bad, so why? Why?' But my parents know I am happy, and for them it's the main thing."

NIGERIA

Rabia Salihu Sa'id

ON A SUNDAY MORNING IN APRIL 2012, Rabia Salihu Sa'id, a 52-year-old professor of atmospheric and space-weather physics at Bayero University in Kano, northern Nigeria, was at home with her six children when she heard some shocking news. Fighters from the militant Islamist group Boko Haram had stormed a packed church on the university campus, killing 16 people and wounding many others, mostly students and academics. Sa'id lost two of her colleagues and friends in the attack, which was part of Boko Haram's campaign to build an Islamic state and stamp out Western education.

Sa'id knew that she could have been one of those killed — but it did not stop her from teaching or doing research, and neither have further Boko Haram attacks. "There's a lot of security at the university now, you can't come in without being searched," she says. "But we are not afraid."

Even before these threats, Sa'id's career was riddled with challenges. A growing number of girls in Nigeria receive an education, but in the north of the country — where Sa'id grew up — they often receive little schooling and get married in their teens, says Sa'id's colleague Babatunde Rabiu, director of the Centre for Atmospheric Research at Nigeria's National Space Research and Development Agency in Anyigba. The country is rated poorly in a 2014 index compiled by the Organisation for Economic Co-operation and Development to measure discrimination against women in areas such as law and social norms.

Within science, physics is still very much considered a 'male' subject, says Rabiu. A report[1] that Sa'id co-authored in 2013 found that 5–20% of students enrolled in physics in Nigerian universities were women, and UNESCO figures show that women make up 24% of doctoral students in all areas of science. "It is common for single women who go for further degrees to be mocked by men and even by some females," says Rabiu; traditionally, women are expected to marry and stay at home.

Like Thamyongkit, Sa'id overcame these hurdles thanks to the encouragement of her family. Her father, an officer in the Nigerian Army, encouraged all of his children to pursue education and hoped that his daughter would become a doctor. Instead, she married straight out of school, at the age of 18. When her first son was born with a club foot, she shelved her higher-education plans. Soon, two more children followed. Aged 29, she decided to go to university at last. She did not have a high enough grade in chemistry to enter medicine, so she opted for physics.

But first, there was the pressing question of how to pay for university. She sold a gold necklace from her bridal dowry to pay for the registration, and she ran a nursery school to cover the rest. "Many women would not be able make it like I did, and would have to give up the dream of a university degree if they have no financial support," she says.

Months after Sa'id began her studies, her fourth child was born with sickle-cell anaemia — an incurable genetic disorder. With two children needing medical care, "It was incredibly challenging for me", recalls Sa'id. "I was really in and out of hospitals." Her extended family and friends helped out, staying with one boy or the other at hospital while she went to lectures.

Sa'id was one of 4 women in a class of 21 who graduated in 1996. She says that a streak of ambition kept her going. "I wanted to be in a position where my opinion will matter, where I will be respected,

MUSTAPHA SHUAIBU HASSAN

Rohini Godbole, a physicist at the Indian Institute of Science in Bangalore.

P. N. BALA SUBRAMANIAN

INDIA

Rohini Godbole

"I HAVE READ YOUR HUSBAND'S PAPERS — great work!" exclaimed a German physicist as he shook hands with Rohini Godbole. The seminal papers on particle physics, he assumed, had been written by a man. But the author was Godbole herself — and the remark was one of many that she has deflected throughout her career.

Based at the Indian Institute of Science in Bangalore, Godbole is one of her country's few leading physicists. In 1991, she discovered a way to describe the complex interplay of high-energy particles in linear colliders[3] with Manuel Drees, now at the University of Bonn in Germany. Godbole is now one of the 16 members of the International Detector Advisory Group for the proposed International Linear Collider.

She has received numerous awards and honours, and is an elected fellow of all three academies of science in India, a nation where about 37% of science PhD-holders, 20% of working scientists and less than 10% of professors are women. Even today, says Anuradha Misra, head of physics at the University of Mumbai, "most promising women scientists in India, in spite of having immense potential, are not able to achieve as much due to societal and cultural constraints".

Godbole was born in Pune, the first scientist in a family rich in doctors and engineers. Her mother was a teacher, and her father said she could be whatever she wanted. She won a coveted government scholarship to study physics, then did a master's at the Indian Institute of Technology (IIT) in Mumbai and a PhD at Stony Brook University in New York. "Until I finished my PhD in physics, I never really thought that being a woman and doing science was somewhat of a special combination," she says.

That soon changed, as Godbole encountered discouragements on every score. When she joined the physics department at the Tata Institute of Fundamental Research (TIFR) in Mumbai for a postdoc in 1979, she found that there was a men's toilet on every floor, but a women's only on every alternate one — because there were so few women working there.

Rabia Salihu Sa'id studies physics at Bayero University in Kano, Nigeria.

where my children could have the good things in life," she says. She went on to get a PhD, then a job as a graduate assistant and eventually a permanent teaching and research post. She now studies the effect of dust aerosols on climate — she has shown, for example, how the constituents of Saharan dust aerosols alter the aerosols' cooling or warming effects[2]. The challenges continue in her research: she has electricity for only four hours each day.

"I wanted to be in a position where my opinion will matter, where I will be respected."

Sa'id and other female scientists in Nigeria have found forms of support. Some research-funding agencies give female applicants particular encouragement or priority, and the Nigerian government has taken steps to reduce discrimination and increase opportunities for women overall. Four years ago, Sa'id and her colleagues formed Nigeria's Association of Women Physicists, to encourage more women in the country into the field. At their first conference, it became clear that all the women had had similar experiences. The group agreed to try to improve representation of women through mentoring, prizes for female students, efforts to improve school physics teaching and a meeting twice a year.

Last month, at the annual meeting of the American Association for the Advancement of Science in San Jose, California, Sa'id received an Elsevier Foundation Award for Women in Science in the Developing World — she was one of five women honoured for work in physics and mathematics. "I have learnt that with high motivation and hard work, one can succeed," she says. But women cannot progress without improvements in the country's research infrastructure as a whole, she says. "This is very important."

In some other physics departments, there were no female toilets at all.

At the end of her postdoc, she was told by a senior professor that she should take up teaching in a women's college. "There's no expectation that you could be at the top, leading something, having a lab, directing PhD students," she says. It was advice she chose to ignore, and soon she had a job as a lecturer at the University of Mumbai.

The subtle — and not so subtle — messages have kept on coming, says Godbole. Five years ago, she was elected a distinguished alumna of IIT, but women were still such a rarity in the group that the letter addressed her as 'Dear Sir'. "These things could have put me down," she says, but her energy carried her through. At Mumbai, she would usually lecture and work at the university until noon, then go to the TIFR some 30 kilometres away to continue her collaborations there. She would pause for dinner, work until 11 p.m., and then be ready to teach again the next day.

The excitement of discovery was a driver too. "Once in a while you figure out something that, until then, others have missed, and you feel suddenly like, 'Hey, I know something that others did not know!'" she says. By 1994, she was a professor at the Indian Institute of Science.

Meanwhile, Godbole's family had been fretting that she was getting old and would never find a husband. Indian society expects women to put marriage and children first, and to weave their job around the family, says Vinita Sharma, former head of Science for Equity, Empowerment and Development, a Ministry of Science and Technology initiative that promotes science projects to help disadvantaged sections of society.

Perceptions and attitudes are changing. In big cities, it is now considered acceptable for women to study physics. The proportion of women starting a bachelor's degree in science rose from 20% in 1970 to 40% in 2005, the most recent figures available. Half of the students doing a physics master's at the University of Mumbai are now women. In 2003, the Ministry of Science and Technology started a programme to encourage female scientists to return to research after a career break. And in 2005, the government set up a National Task Force for Women in Science, which made recommendations, but has had limited impact.

Shobhana Narasimhan, a physicist at the Jawaharlal Nehru Centre for Advanced Scientific Research in Bangalore, was on the task force. She says that there are two issues. First, the need to raise women's aspirations — for this, leadership programmes can help, she says. (Godbole chairs a panel to study the issue of women in science, and has undertaken several initiatives, such as mentoring aspiring female scientists.) The second is the need to change the attitudes of men, so that they consider women for leadership positions. "This is hard!" says Narasimhan. "I think it helps for women to just 'hang out' with the guys and get accepted as one of them, but in conservative societies, this can often cause a lot of problems."

Godbole did get married, but never had children. "I didn't see how I could juggle all the balls," she says. "If there had been child care at institutions, it would have tipped the balance in terms of my decision not to have children. That's the only regret I have."

She thinks future scientists should not be forced to make that choice, and she would like to see better child-care provision and extensions to postdoctoral contracts for women who have children during their positions, as is commonly done in Europe.

Murenzi points to the significant cultural evolution that has opened up opportunities for women in places such as the United States and Europe: there is much more awareness of the obstacles than there once was, and greater efforts to minimize them. "For the most part, developing nations have not come so far, so quickly," he says. "Make no mistake, though — that evolution definitely is under way in many developing countries."

A few years ago, Godbole co-edited a book of biographies of almost 90 female scientists in India[4], in which she wrote that a female scientist needs a large dose of luck to have a successful career, a happy family and a happy marriage. Today, she says: "Whatever can be done to take the requirement of luck from this equation will be good." ■

Katia Moskvitch *is a science writer in London, and was an International Development Research Centre fellow at* Nature.

1. Fuwape, I., Fasunwon, S., Obiekezie, N. & Said, R. *AIP Conf. Proc.* **1517,** 132 (2013).
2. Sa'id, R. S. & Rashmibila, M. M. *J. Niger. Assoc. Math. Phys.* **20,** 331–342 (2012).
3. Drees, M. & Godbole, R. M. *Phys. Rev. Lett.* **67,** 1189 (1991).
4. Godbole, R. & Ramaswamy, R. (eds) *Lilavati's Daughters* (Indian Acad. Sci., 2008).

Uninfected but affected: women line up for perinatal care at a Marie Stopes centre in Sierra Leone.

Ebola's lasting legacy

One of the most devastating consequences of the Ebola outbreak will be its impact on maternal health.

BY ERIKA CHECK HAYDEN

The woman was lying still on the floor of the ambulance when Emma Åkerlund opened the doors. Between her legs was a tangle of fabric with something inside — the body of a newborn boy with the placenta and umbilical cord twisted up in his mother's skirt. If he survived the delivery, he did not last long after.

The woman had gone into labour earlier that day in November last year. But health workers feared that she had Ebola, and refused to treat her. She had spent hours alone in the back of the ambulance as it made its way to ELWA3, an Ebola management centre in Monrovia, Liberia, run by the medical aid group Médecins Sans Frontières (MSF).

Åkerlund, a gynaecologist and obstetrician, was able to examine the woman while wearing full protective gear: a waterproof body suit, gloves, boots and goggles. She asked the woman if she wanted to see her son; the mother declined. So Åkerlund zipped the baby into a body bag and sterilized it with bleach, just in case the mother and child were infected.

Tests performed later that day proved that they were not. There had been no medical reason for doctors and nurses to turn the woman away. And although there is no guarantee that the child would have survived with proper medical care, there was no reason for him to come into the world — and then exit it — on the floor of an ambulance.

"I'm not blaming anyone for not taking them in," Åkerlund says. "It's a huge problem to try to take care of pregnant women in an Ebola epidemic."

Ebola is having tremendous knock-on effects for maternal health in Liberia, Guinea and Sierra Leone. Pregnancy seems to make women uniquely vulnerable to the effects of the disease, and babies born to infected women have not been known to survive. Compounding these individual tragedies, the blood and abundant bodily fluid that accompanies delivery or miscarriage pose enormous risk of infection to health workers. As a result, many refuse to treat patients who are pregnant for fear that they will become infected. And throughout the region, fears about Ebola and stories about women being turned away have convinced many pregnant women to stop showing up for routine prenatal visits or for assistance with delivery.

The epidemic has so far infected more than 23,900 people, killing 9,700 of those, and although it is on the wane it is unclear when it will end. For maternal health, the effects are devastating. The United Nations Population Fund (UNFPA) estimates that, either directly or indirectly, the epidemic will result in as many 120,000 maternal deaths by the end of October.

Researchers, public-health experts and activists worry that this trend could undermine advances made in the region's health, education and more. For instance, children whose mothers die may end up orphans and be forced to choose work over school. Teenagers have found themselves becoming family breadwinners.

"The impact on maternal and child health will be one of the most important and long lasting from this outbreak," says Alexandre Delamou, chief of research at Guinea's National Centre for Training and Research in Maferinyah.

FEARSOME ODDS

Doctors are still trying to work out why Ebola is more deadly for some people than others. For pregnant women, the consequences of infection seem dire. In the first Ebola outbreak ever identified, in Zaire in 1976, 73 of 82 infected pregnant women died. In a 1995 outbreak in Kikwit, Democratic Republic of the Congo, researchers observed that all but one of 15 pregnant women treated at one hospital died.

Such samples are too small to say anything with certainty, but the idea that pregnant women would be more at risk of dying makes sense. Ebola increases the risk of haemorrhage, which is already elevated in childbirth. And pregnant women are more likely than other adults to die from several other infectious diseases because of a natural suppression of immunity that helps to protect developing fetuses.

The fetuses never seem to make it. In every case reported in the literature of a woman with Ebola giving birth, the baby has either been stillborn or died shortly thereafter. Why this happens is unclear; it could be because of the high viral load that crosses the placenta, because the virus is transmitted through breast milk or because mothers who are sick or recovering from Ebola are in no shape to care for newborn babies.

Pregnant women with Ebola also pose serious risks to caregivers. The symptoms of Ebola for a pregnant woman — abdominal pain, vaginal bleeding, premature labour, nausea, diarrhoea and vomiting — can be very difficult to differentiate from standard pregnancy complications. Given the high exposure to bodily fluids during delivery, those who guess wrong often do so at their peril.

> *The Ebola epidemic may cause as many as* **120,000 maternal deaths** *by the end of October.*

About one-third of the medical staff who died from Ebola in Sierra Leone between April and September last year, for instance, were mother-and-child health aides. This group had a higher death rate than other types of health worker, such as doctors or nurses.

Amadu Jawara, a nurse and community health worker in Freetown, Sierra Leone, is one of the lucky ones. In November, he and Samuel Batty treated a pregnant woman with abdominal pain and a fever. Because the fever responded to medication, they assumed she had malaria and did not use full protective gear while examining her. But she soon worsened and later died of Ebola. Jawara and Batty were placed in isolation for 21 days. Batty died from the disease on 2 December; Jawara did not get infected.

On the day he reported back to work, six women arrived at the hospital in need of Caesarean sections. Many doctors and nurses had abandoned their posts during the epidemic, and there were no surgeons available. Jawara, who had been trained in basic surgery, was the only person there to perform the deliveries. "They told me I should intervene," Jawara says. "I had no choice." He has since performed 100 such procedures.

The staff try to screen out patients who have a fever or demand that they take an Ebola test. But this leaves some women without care during labour, and some deliver on the floor of the waiting room with no one to help them. And screening is not foolproof.

On 13 January, Jawara saw a patient who had been four months pregnant but whose fetus had stopped moving. She had a low-grade fever, and Jawara admitted her, believing that she had an infection.

Two days later, she expelled her fetus and Jawara helped to deliver her placenta. The woman soon developed diarrhoea and a high fever. She tested positive for Ebola.

Jawara's lucky streak continued; he did not become infected. But every episode like this deepens the fear that health-care workers have of treating pregnant women. Some worry that the rift between the two will outlast the epidemic. "Ebola will have an everlasting impact on the health system in Sierra Leone," Jawara says. "The fear is still there." Even as the epidemic winds down, many health workers remain absent from their posts.

TURNED AWAY

Researchers who study maternal mortality talk about the "three delays" that increase the proportion of women who die in childbirth: delays in deciding to go to a hospital, delays in getting there and delays after arriving caused by factors such as electricity cuts or the lack of a trained doctor. Ebola has exacerbated all three.

In September, a team of researchers led by the US Centers for Disease Control and Prevention (CDC), for example, interviewed pregnant women and health-care workers in Kenema, a town in southeastern Sierra Leone that was at the epicentre of the epidemic when it started in the country. They found that pregnant women stayed away from hospitals for fear of contracting Ebola, or because of rumours that health-care staff were injecting patients with Ebola, harvesting their blood for sale or misdiagnosing them with Ebola because they were paid extra for referring them to treatment units. Even after outreach workers conducted education campaigns in local communities, all of the 27 pregnant or breastfeeding women who participated in the study said that they knew of at least one person who was refusing to go to health-care facilities because of such rumours.

Many women who wanted to go to clinics were prevented from reaching them by quarantines, roadblocks or a lack of transportation. And those who made it were often turned away because the staff were afraid of getting infected. Many hospitals were formally or effectively closed because staff stopped going to work. In Freetown, the capital and largest city of Sierra Leone, the biggest public maternity hospital has been running on skeleton staff for much of the outbreak. In the second largest city, Bo, MSF closed one of the few programmes in the country that offered advanced surgical care for pregnancy complications last August. The decision came after several instances in which staff had difficulty diagnosing Ebola cases. It has still not reopened.

CDC researchers found from their surveys

that education campaigns are helping to ease fears in regions of the epidemic that are no longer seeing many new cases. The immediate damage is nevertheless erasing fragile gains that these countries had made in maternal health.

LYNSEY ADDARIO/VII/MSF

Amadu Jawara (right) does rounds at a medical facility in Bo, Sierra Leone, in 2012.

DELIVERING CHANGE

Before the outbreak began, Liberia, Guinea and Sierra Leone had some of the world's highest maternal death rates: in Sierra Leone, more than 1,000 of every 100,000 child births results in the death of the mother. That is hundreds of times the ratio for developed countries in Europe. But it actually represented a vast improvement from where Sierra Leone was near the end of its civil war in 2000, when the death rate was twice as high. Rates of death during childbirth had also dropped in Liberia and Guinea thanks to a combination of national governments declaring maternal mortality an important health priority, an influx of foreign aid and major public-health campaigns. The gains have been made despite a lack of trained medical professionals and the fact that these countries are some of the poorest on Earth.

"It's like we had taken one step forward before Ebola and now we've taken three steps back," says Shumon Sengupta, country director in Sierra Leone for Marie Stopes, a non-governmental organization that provides family-planning services. A study conducted on behalf of the UNFPA, for instance, modelled the impact of the reductions on routine care provided to pregnant women in Sierra Leone in the first six months of the epidemic. The preliminary work found that the cuts may have caused 20% more maternal deaths.

Now, the maternal death rate is projected to double to more than 1,000 per 100,000 live births in Guinea and Liberia and to more than 2,000 per 100,000 live births in Sierra Leone, returning to wartime levels.

"We think the collateral damage of the epidemic is higher than the damage caused by the epidemic itself," says gynaecologist Séverine Caluwaerts, who works at the Institute of Tropical Medicine in Antwerp, Belgium, and with MSF.

The UNFPA is looking to convene a meeting to analyse the state of the science around reproductive health and Ebola and how to prevent such impacts during future outbreaks. A crucial issue is how to boost the survival of pregnant women who become infected.

An estimated **1.2 million women** *may lack access to family-planning services, which will increase the number of unexpected pregnancies.*

MSF obstetrician Benjamin Black and his colleagues have designed a 'maternity box' containing all the supplies that caregivers would need to help prevent haemorrhage in cases in which the uterus fails to contract after delivery, a major complication that threatens pregnant women with Ebola. These supplies include drugs such as oxytocin and misoprostol, which encourage contraction, and a catheter to empty the bladder, which can speed up delivery.

The box spares caregivers from having to exit the patient area, take off their protective equipment and re-gown if they need something vital. Simple adjustments such as this have helped MSF to save the lives of 23 pregnant women with Ebola.

In January, MSF opened a clinic in the outskirts of Freetown that specializes in care for pregnant women with Ebola. The clinic has adapted its procedures to meet the additional risks in looking after these women — for instance, adding a third pair of gloves that reach the elbow to the standard two pairs. And staff try to avoid surgical procedures, such as episiotomies and Caesarean sections, that are too dangerous for workers and patients.

Åkerlund says that she is constantly thinking about how to protect herself and other staff so that they feel able to care for pregnant women rather than turning them away.

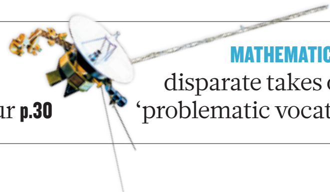
One case from her time in ELWA3 haunts her: an apparently full-term pregnant woman who arrived at the treatment centre without a single symptom of Ebola. She had thought that her waters had broken and she headed for the hospital but was turned away from multiple clinics by reticent staff. Åkerlund thought that the woman was another perfectly healthy mother-to-be who might lose her baby to nothing more than fear. She and her staff admitted the woman, brought her to a place where she could be kept separate from other patients and ordered an Ebola test.

Hours later, Åkerlund received the results: the test was positive. A second test was also positive. The woman was moved into the ward for confirmed cases on 26 November. Within days, her condition deteriorated; her blood stopped clotting, and her baby stopped moving. The following day, the woman died.

The case sticks in Åkerlund's mind as an example of the dilemmas that confront both caregivers and pregnant women in an Ebola epidemic, when diagnostic tests can take hours or days and a woman in labour may need help immediately. But Åkerlund knows that giving up is not an option. "I don't think the solution is to say you can't do maternity care in an outbreak and leave these women without anything." ■ SEE EDITORIAL P.5

Erika Check Hayden *writes for* Nature *from San Francisco. The Pulitzer Center on Crisis Reporting provided support for this coverage.*

COMMENT



SPACE Voyager memoir hymns the little probes that went on a grand tour **p.30**

MATHEMATICS Two disparate takes on a ‘problematic vocation’ **p.31**

DOCUMENTARY How aid after the 2004 tsunami wrought a second disaster **p.32**

OBITUARY Carl Djerassi, pill pioneer, novelist and playwright, remembered **p.34**

ILLUSTRATIONS BY DEREK BACON

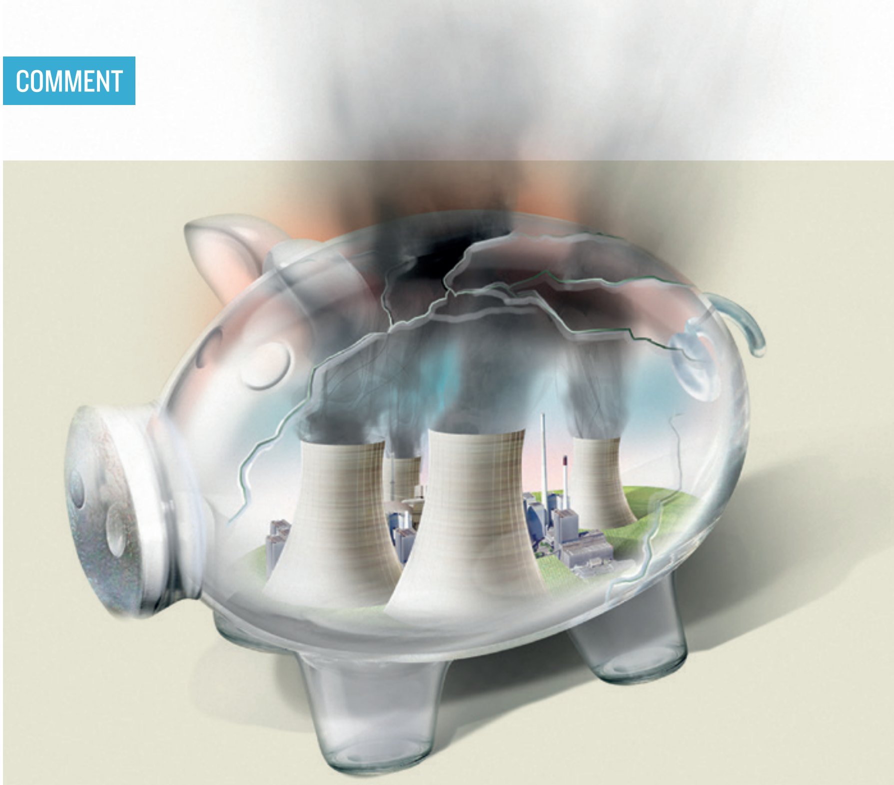
Support low-carbon investment

Private finance can drive the energy transformation needed to meet global emissions goals — if backed by the right policies, says **Nathan Fabian**.

Markets and governments are converging to address climate change. As scientific evidence and government actions strengthen, investors and financiers are reducing the exposure of their portfolios to risks from rising greenhouse-gas emissions. They are allocating more capital to low-carbon activities and less to carbon-intensive industries[1].

In September 2014, banks, insurance companies, charities, and pension, mutual and endowment funds announced that they would direct an extra US$125 billion per year until 2020 to investments that address climate change. Fossil fuels are being divested from by influential funds, including the Rockefeller Brothers Fund of New York, and universities in the United States, the United Kingdom and Australia.

Those commitments represent just the tip of the financial iceberg. Around $300-trillion worth of assets is managed globally[2] and more than $20 trillion of new investment is forecast to flow into the global economy each year[3]. Private finance has the capacity to fund the wholesale shift to a low-carbon economy that the world needs to keep global warming to within 2 °C, the limit agreed on by national governments. The International Energy Agency estimates that such a ▶



FINANCE SHIFTS

Black to green

Low-carbon investment announcements totalling more than US$600 billion at the 2014 United Nations climate summit in New York highlight how the market is moving.

Insurance companies pledged to invest $420 billion over five years in renewable energy, energy efficiency and sustainable agriculture projects.

Commercial banks will issue $30 billion in green bonds in 2015, building on $37-billion worth issued in 2014.

Foundations are abandoning fossil fuels. The Rockefeller Brothers Fund, an investment fund based on oil wealth, led a $50-billion coalition that will divest from fossil fuels in the next three to five years.

University endowments are shifting capital out of emissions-intensive activities. Stanford University in California is selling its shares in coal companies; Britain's Glasgow University is freezing new fossil-fuel investments; and in February 2015, the University of Sydney in Australia announced a three-year plan to cut the carbon footprint of its investments by 20%.

Pension funds are following suit. The four regional investor groups on climate change (for Europe, North America, Australia and New Zealand, and Asia) asked fossil-fuel companies in 2014 to justify expenditure on expanding reserves. The pension funds CalSTRS, APG and PensionDanmark committed up to $31 billion in low-carbon allocations by 2020. The Portfolio Decarbonization Coalition, a group of pension funds and fund managers established in September 2014 and led by the UN Environment Programme Finance Initiative, is encouraging investors to decrease the carbon exposure of $100 billion of assets by the end of 2015 (see go.nature.com/fczzfd). **N.F.**

▶ transformation could be achieved by a sixfold increase[4] in annual investments in clean energy and energy efficiency, from around $390 billion in 2013 to $2.3 trillion per year by 2035 (see 'Investors clean up').

Government policies are the sticking points. Carbon pricing, favourable tax regimes and rebalancing of subsidies will encourage demand for low-carbon capital. Supply will be increased by government support for research and development and improved financial policies. Countries that understand these dynamics will benefit from more private financing and clean growth, as are Vietnam and Germany now. Those that do not will be left squeezing the last remnants of value out of waning industries.

CASH FLOW

Putting portfolios on a lower carbon footing is in the interests of financiers. In exchange for providing capital to support economic activity, investors earn a return, at a rate that reflects the level of risk. Because of the growing costs of emissions, low-carbon activities such as renewable energy and green buildings are a safer bet than many carbon-intensive ones. The prospect of unburnable fossil-fuel reserves makes many new coal and tar-sands projects too risky to invest in.

Policies can also make low-carbon investments profitable. For example, tax credits are provided for renewable-energy projects in the United States. Feed-in tariffs (payments for energy supplied to the grid) are offered in Germany, Malaysia and South Korea. Carbon-pricing schemes now cover 22% of global emissions. Public institutions such as the World Bank and the European Investment Bank are issuing green bonds in record volumes (see 'Green bonds'), and there are more than ten government-funded green investment banks around the world that co-invest with the private sector to reduce risk and improve financial returns. By contrast, fossil-fuel commodity markets are volatile and threatened by political instability.

MINIMIZING RISK

Banks and investors increasingly see climate change as a raft of risks that they must respond to directly rather than waiting for governments alone to fix (see 'Black to green'). For example, by early 2015, nine international banks had chosen not to invest in the development of the Carmichael coal mine in Queensland, Australia, which is part of one of the largest coal basins in the world. Leading financiers are also under public pressure to manage risk more carefully. Former US Secretary of the Treasury Hank Paulson wrote last year[5] that he sees the emerging "climate bubble" as an equally enormous danger to the world economy as the 2008 financial crisis.

Financial regulators too are indicating that short-term returns should no longer come at the expense of a stable economy and financial system. In June 2014, the UK Law Commission clarified that trustees should account for any factor that is relevant to the financial performance of an investment[6], including environmental impacts.

Better risk analysis from scientists is needed to reinforce policies and accelerate the reallocation of capital. It is especially important to understand non-linear forces such as carbon-cycle feedbacks that could influence global warming, as well as information about rare but severe impacts such as extreme weather events, even at low levels of warming.

How should the low-carbon economy be financed? Robust banking systems, transparent governance and stable currencies are baseline requirements. Predictable long-term energy policies and emissions-reduction frameworks are an essential overlay. If any of these aspects are missing, publicly funded measures can assist. Governments can also raise awareness, build partnerships and change financial regulations to increase private finance flows.

The finance sector's priorities for climate policy were outlined in a statement signed

by around 350 institutional investors contributing more than $24 trillion, presented at the United Nations climate summit in New York last September and delivered to the Group of 20 (G20) governments ahead of its November 2014 meeting in Brisbane, Australia (see investorsonclimatechange.org). The priorities are: carbon pricing, targets or subsidies for renewable-energy and low-carbon technologies, phasing out fossil-fuel subsidies and addressing unintended constraints on the financial system, such as policies that hamper information flows or reduce the availability of investable capital. The United Kingdom, Brazil and the state of California have brought investors into their climate and energy policy-making circles to get the details right.

Carbon pricing is a tool that markets understand. Emissions-trading schemes, carbon taxes or any other mechanism that applies a cost to releasing greenhouse gases are the best ways to distribute the burden of emissions reductions, encourage the market to innovate and achieve abatement cheaply. With lessons learned from the European Union's Emissions Trading Scheme, 73 countries supported the World Bank's renewed push on carbon pricing at the New York summit. Regulation and subsidies — such as fuel standards for cars or emissions standards in energy generation — are not as efficient, flexible or scalable. That said, they have a complementary role and can help to offset existing fossil-fuel subsidies.

Yet governments are vacillating. Spain and Italy have rowed back on emissions reduction and green-technology commitments. Energy feed-in tariffs and renewable-energy targets remain under threat in Australia, whose emissions-trading scheme was repealed in 2014. Financiers now question the economic credentials of such governments. Convinced of the benefits of low-carbon investments, they will increasingly take their money elsewhere.

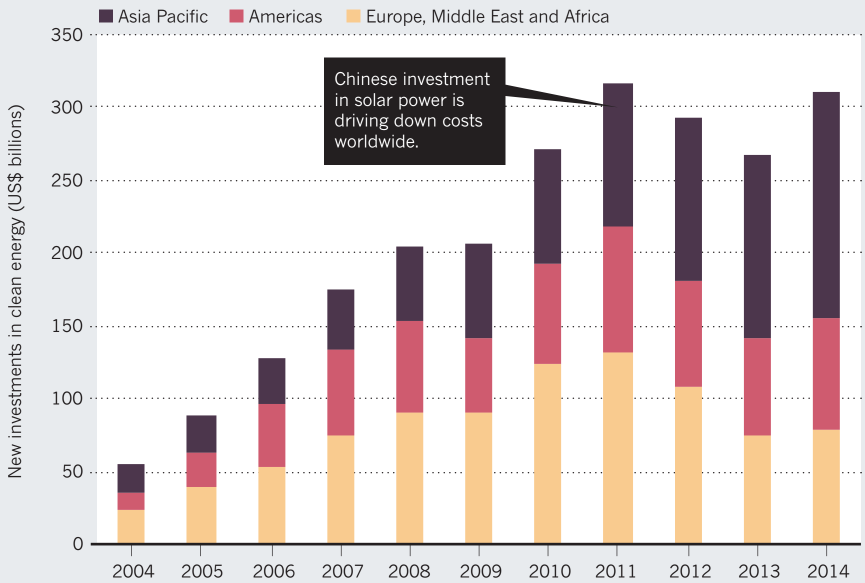
Capital is mobile. Pension funds, asset managers and insurance companies allocate up to 50% of their assets to international markets. Even though cross-border, low-carbon financial flows have been relatively low[7], they will grow as policies improve, investor experience increases, trade costs come down and new opportunities emerge. Investors will shift more capital to the offshore markets that best meet their long-term objectives.

NEGOTIATING TABLE

The role of private finance should be taken more seriously in the international climate-negotiation process. Political posturing on the relative levels of emissions cuts is too far removed from the workings of the real economy. As well as negotiating with each other, countries should pitch their emissions-reductions plans to the finance community for input and capital at international meetings such as the G20 summits.

INVESTORS CLEAN UP

Flows of new finance into renewables, energy storage and distribution have increased more than fivefold in the past decade. Since 2012, funding has grown in China but has fallen in Europe.



SOURCE: REF.1

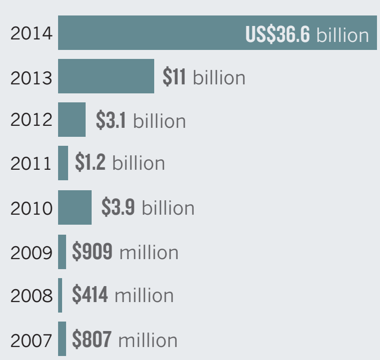
Finance flows to less developed and emerging economies will help to achieve the 2 °C goal. Multilateral and national development banks must continue to expand their roles. Investment funds and financial instruments that reduce currency volatility, select the best projects and partially absorb unexpected losses can increase financial flows. The Private Sector Facility of the Green Climate Fund — a component of the UN Framework Convention on Climate Change — will play an important part in leveraging private finance, increasing the size and impact of public climate finance.

Even with the best of intentions, policy missteps will occur. Financiers should see through short-term policy volatility by backing governments with clear low-carbon policy agendas and making them aware of the consequences of unpredictable changes. Because markets cannot solve climate change alone, authorities such as the Financial Stability Board, which monitors the global financial system, must continue to provide direction.

If appropriate policies are delivered, the capital shift that is under way will gather steam. If not, investment will not be enough to meet the 2 °C global-warming limit. ■

GREEN BONDS

Banks, governments and companies have issued tens of billions of dollars' worth of bonds to finance sustainable projects. Investors earn regular interest income and a further payment when the bond matures after a fixed period.

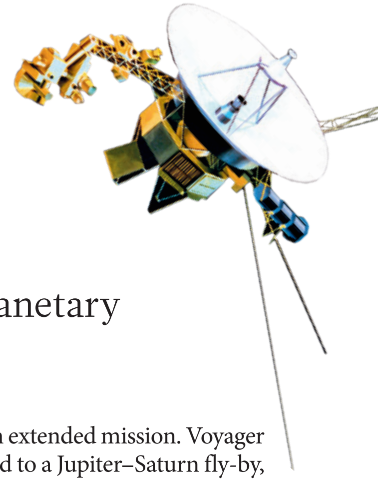


SOURCE: CLIMATE BONDS INITIATIVE

Nathan Fabian *is chief executive of the Investor Group on Climate Change Australia and New Zealand in Sydney, Australia.*
e-mail: nathan.fabian@igcc.org.au

1. Mills, L. *Global Trends in Clean Energy Investment* (Bloomberg New Energy Finance, 2015).
2. United Nations Environment Programme. *Aligning the Financial System with Sustainable Development: Pathways to Scale* (UNEP, 2015).
3. The Global Commission on the Economy and Climate. *The New Climate Economy Report* (The Global Commission on the Economy and Climate, 2014).
4. International Energy Agency. *Special Report: World Energy Investment Outlook* (IEA, 2014).
5. Paulson, H. M. Jr 'The coming climate crash' *The New York Times* (21 June 2014).
6. UK Law Commission. *Fiduciary Duties of Investment Intermediaries* (UK Law Commission, 2014).
7. Organisation for Economic Co-operation and Development. *Mapping Channels to Mobilise Institutional Investment in Sustainable Energy* (OECD, 2015).

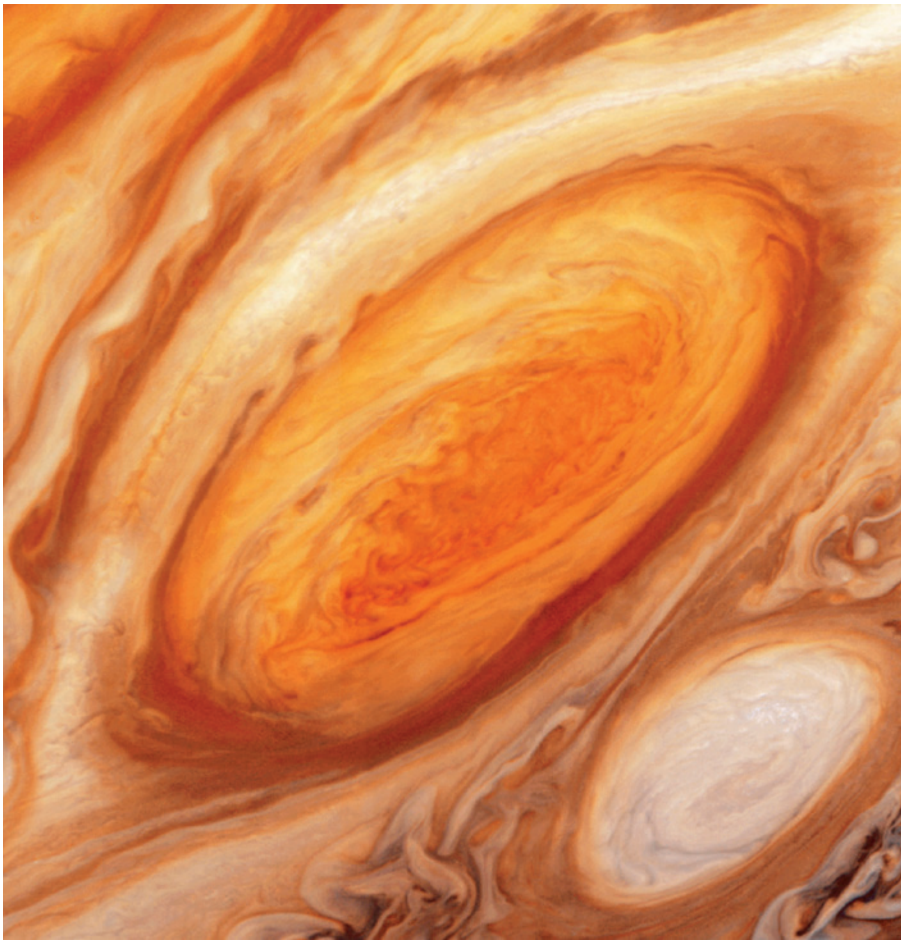
The author declares competing financial interests: see go.nature.com/xmglaw for details.



SPACE SCIENCE

Beyond the heliopause

Roger D. Launius savours a masterful account, by a veteran of interplanetary space science, of the Voyager probes' mission to the giant planets.



NASA/SPL

Jupiter's Great Red Spot, a massive stable storm photographed by Voyager 2 in 1979.

Part memoir, part anecdotal history and part sermon on the delights of science, *The Interstellar Age* is a captivating read. In it, planetary scientist Jim Bell presents the eventful story of NASA's Voyager 1 and Voyager 2 missions to the edge of the Solar System and beyond. Bell, a veteran of many space-science missions, including several Mars probes, brings deft writing and an in-depth, nuanced understanding of big planetary-science efforts to this popular account.

Voyager's legendary status has been long assured, although *The Interstellar Age* will add to its cachet. Conceived in the 1960s and launched in the 1970s, the twin probes encountered the larger outer planets of the Solar System — Jupiter, Saturn, Uranus and Neptune — between the late 1970s and the late 1980s. Having gone far beyond their original remit, they now continue an interstellar mission at the edge of the Solar System. The probes are, in essence, 'the little spacecraft that could'.

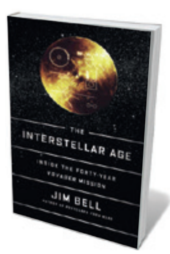
Bell writes about how, in the early 1960s, aerospace engineer Gary Flandro and other scientists realized that once every 176 years, Earth and the Solar System's four giant planets gather on one side of the Sun. This would enable close-up observation of the giants, in a planetary 'grand tour'. During the fly-by of each planet, a gravity assist — a slingshot effect, harnessing the planet's movement and gravity — could increase the spacecraft's velocity and reduce flight times.

Bell describes the politics of the planning stage. The four-planet scenario was possible, but NASA deemed it too expensive to build a spacecraft for an extended mission. Voyager was downgraded to a Jupiter–Saturn fly-by, but engineers designed as much longevity into the probes as the US$875-million budget would allow. NASA launched Voyager 2 on 20 August 1977, and Voyager 1 followed on a faster, shorter trajectory on 5 September.

Bell, who hung out with the science team as an undergraduate at the California Institute of Technology in Pasadena, describes how the craft achieved their objectives — and then some. Voyager 1 was, for example, programmed for a close encounter with Saturn's moon Titan, during which it revealed a complex world with an atmosphere, thick clouds and water ice. It showed that Titan was ripe for scientific investigation, paving the way for the sustained investigations of the Huygens–Cassini mission at the dawn of this century. But this fly-by deflected Voyager 1 out of the Solar System's elliptical plane; unable to continue on to Uranus and Neptune, the craft's planetary mission ended.

Voyager 2 continued on to the two outermost giant planets. Bell reports how, as the probes flew, controllers constantly reprogrammed the on-board computers, which had only about 5,000 words' worth of memory each, to take advantage of scientific opportunities. Successfully capturing data was hugely taxing, but mission engineers and scientists made it work.

The probes explored the giant planets' systems of rings and magnetic fields, finding previously unknown geological activity on Io, a moon of Jupiter with numerous volcanic features. The Voyagers also explored a total of 48 moons around the gas giants. They sent back more than 100,000 images of the planets, rings and satellites, and took magnetic measurements, chemical spectra and radiation readings. The information revolutionized Solar System science, helping to resolve key questions about how it was formed and raising intriguing new ideas such as the possibilities of life beyond Earth.



The Interstellar Age: Inside the Forty-Year Voyager Mission
JIM BELL
Dutton: 2015.

Bell is at his best in telling the human

The identical Voyager probes launched in 1977 and are still travelling.

NASA/JPL/SPL

stories of discovery, excitement and public engagement. He describes the extension of the Voyager mission to the heliopause, where the Sun's energy is overpowered by interstellar forces. Ed Stone, who has been chief scientist for Voyager since its inception, evinced the excitement of a self-confessed non-party animal. "I can still remember taking the data home every night, and putting the plots on the refrigerator," he tells Bell. "I couldn't stop thinking about them, wondering what would happen next."

Voyager 1 is now more than 130 astronomical units (AU) from the Sun (1 AU is the distance between the Sun and Earth, around 150 million kilometres), and Voyager 2 is at more than 107 AU. They continue to take readings of the heliopause.

Thanks to astronomer Carl Sagan, one of Bell's heroes, both probes contain messages from Earth: gold-plated copper phonograph records encoded with 115 images of scenes from Earth, audio greetings in 55 languages, and 90 minutes of music from Bach to Chuck Berry, along with playing instructions. This message in a bottle is one of the mission's best-known attributes, and Bell explains well its publicity value and how it represents a feel-good sentiment about the possibility of encountering interstellar life.

The Voyagers demonstrate the remarkable advances in robotic space exploration over the past almost 40 years, and suggest that subsequent missions may yield even more exciting results. Such follow-ups as the Galileo mission to Jupiter, Huygens–Cassini to Saturn and New Horizons to the Kuiper belt including Pluto may herald even more ambitious missions — to Titan, for instance, where they might sail on a hydrocarbon sea, or to Jupiter's moon Europa, to explore an ice-covered liquid-water ocean that has the potential to harbour life.

I believe that NASA's greatest achievement is the Apollo Moon programme. The odyssey of the Voyagers certainly vies for second place. Bell appropriately quotes historian Stephen Pyne: "The Voyagers were special when they launched. They have become more so thanks to their longevity, the breadth of their discoveries, the cultural payload they carried, and the sheer audacity of their quest." ■

Roger D. Launius *is associate director for collections and curatorial affairs at the Smithsonian Institution's National Air and Space Museum in Washington DC.*
e-mail: launiusr@si.edu

MATHEMATICS

Groping in the dark for glimpses of beauty

Amir Alexander relishes two accomplished accounts of the life mathematical.

Birth of a Theorem: A Mathematical Adventure
CÉDRIC VILLANI
Bodley Head/Faber and Faber: 2015.

Mathematics Without Apologies: Portrait of a Problematic Vocation
MICHAEL HARRIS
Princeton Univ. Press: 2015.

One evening in 2009, French mathematician Cédric Villani stepped into his children's room. He locked the door, turned off the light and began to pace, pondering the statistical properties of plasma. A few metres away, his wife, Claire, was in the kitchen, cooking dinner for the family. The contrast, Villani concedes in his engaging *Birth of a Theorem*, was "a bit much", yet immediately after dinner he returned to the dark room to grapple for hours with his elusive proof.

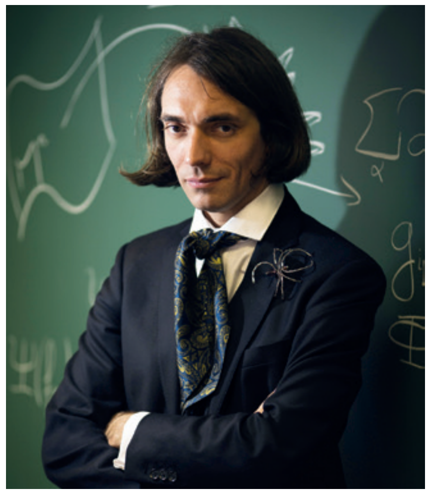
Anyone reading that anecdote will feel sympathy for Claire as she tries to preserve family normality. Anyone who has seriously engaged with mathematics will also understand her husband. Perhaps more than any other field, mathematics pulls the practitioner away from the 'normal' world of things and people into a strange alternate universe, in which we catch glimpses of beauty and coherence, but spend most of our time groping in the dark. In *Birth of a Theorem*, Villani offers one way of straddling that divide; in *Mathematics Without Apologies*, fellow mathematician Michael Harris presents a very different one. Together, they provide an unmatched perspective on life in this "problematic vocation" by two of its leading practitioners.

Birth of a Theorem is the story of Villani's quest to give a full mathematical account of Landau damping. Whereas gas becomes increasingly disordered over time as entropy increases, plasma spontaneously stabilizes, with no increase in entropy. Soviet physicist Lev Landau was the first to mathematically describe this improbable phenomenon, but he used a simplified model that left many unconvinced. Working closely for several years with mathematician Clément Mohout, Villani succeeded, and he was awarded a Fields Medal in 2010.

Villani's quest takes him across the world, from Lyons, France, to Princeton, back to Paris and on to Hyderabad in India. At every stop, he talks to local mathematicians, demonstrating that, for all its abstractness, mathematics can be an intensely social activity. The book is sprinkled with brief, telling

➲ NATURE.COM
For more on science in culture, see:
nature.com/booksandarts

JOEL SAGET/AFP/GETTY



Cédric Villani studies the maths of plasmas.

portrayals of mathematicians and physicists past and present. The grumpy, grey-haired Étienne Ghys of Lyons and Chinese expat Alice Chang of Princeton alternate with the autocratic Lev Landau in 1960s Moscow and the ever-present shadow of Albert Einstein at the Institute for Advanced Study in Princeton. Much of Villani's e-mail correspondence with Mohout is reproduced, chronicling moments of triumph and despair.

Charismatic and flamboyantly dressed, Villani is the opposite of the 'mathematical hermit' and annoyed by the stereotype. He attends a recital by one of his children, joins his family at the American Museum of Natural History in New York and travels a long way to attend a concert by rock band Têtes Raides. Yet he studies the mathematics of galaxy formation during the recital, works out a step in his proof on the bus from the museum, and explains his research to a stranger who drives him back from the concert. The mathematical life, in his telling, is a delicate dance between the demands of the 'real' world and the ▶

allure of the mathematical one.

If *Birth of a Theorem* is the personal record of a single-minded quest, *Mathematics Without Apologies* is a kaleidoscope of philosophical, sociological, historical and literary perspectives on what mathematicians do, and why. Do they pursue their work for the public good? Harris dismisses that as a pose, useful for grant applications and little else. Is it the absolute truth of mathematical demonstrations that drives the field? That, Harris contends, is a conceit of philosophers: practising mathematicians seek insight, not certainty. What about the lauded beauty of mathematics? Perhaps, Harris concedes, but when mathematicians talk about beauty, what they mean is pleasure. A 2012 sociological survey found that 91% of pure mathematicians cited it as a key attribute of the field. Mathematicians, Harris concludes, do what they do because of the enormous pleasure it brings them.

Pleasure is not an explanation likely to satisfy funding agencies. Yet Harris makes no apologies. He is concerned that the field has made a Faustian bargain with the institutions that provide the material conditions for mathematical research. Leading mathematics departments have trained an army of quantitative analysts, or quants, to implement the algorithms that govern financial trading practices. The impersonal, unchallengeable equations of higher mathematics, he worries, contribute to a moral vacuum at the heart of high finance.

Harris's insider view reveals a community in which each mathematician is placed in an informal but strict hierarchy, depending on acknowledged brilliance and accomplishments. He takes a playful detour, arguing that each of US writer Thomas Pynchon's 'non-linear' novels is organized around a different conic section, such as a parabola. Throughout the book, he verbally spars with an imaginary "performing artist" while trying to explain the mysteries of number theory.

But, like Villani, Harris returns repeatedly to the chasm between the human world and the mathematical one — a tension that in his own life has proved fruitful. Stuck in a professional cul-de-sac in the 1990s, Harris experienced a revelation: a dream showed him a new mathematical path, which led to his ascent up the mathematical hierarchy and transformed his life.

For him as for Villani, mathematical insight at its deepest core remains an irreducible personal experience. ■

Amir Alexander *is the author, most recently, of* Infinitesimal. *He teaches at the University of California, Los Angeles.*
e-mail: amiralex@ucla.edu

SOCIAL SCIENCE

Aid's inconvenient truth

Erin Bohensky applauds a documentary revealing how disaster relief can have disastrous impacts.

At the start of Raphael Barth's provocative documentary *Aftermath*, a bottle of Coca-Cola lands on a pristine beach. The image calls to mind the satirical 1980 film *The Gods Must Be Crazy*, in which the lives of Kalahari Desert tribal peoples are changed irreversibly by modernization. Barth's is the true story of how foreign aid delivered to inhabitants of the Nicobar Islands after the 2004 Indian Ocean tsunami became a second disaster.

Barth's film focuses on social ecologist Simron Singh and his work with these indigenous peoples. In setting out to help them, Singh grapples with two questions: why is aid so dysfunctional, and how can science help people in crisis?

Singh first visited the islands in 1999 to research Nicobarese culture, and befriended many in the community. After the tsunami, they asked him for help. Singh mobilized funding from the Austrian Science Fund to support rehabilitation research. In the film, he — with community spokesperson Prince Rasheed Yusoof and local reporter Denis Giles — watches aid pour in from hundreds of non-governmental organizations (NGOs). In effect, the tsunami engineers the perfect social experiment, revealing what happens when cash and unneeded commodities are funnelled to a remote indigenous community. There was "no way back", Singh admits.

"Even well-directed intervention is no substitute for empowerment."

The donations, Barth shows, are mismatched to needs and context. The community wants tools; donors bring blankets. Houses and schools rebuilt with aid are ill-ventilated "boiling chambers". A government initiative provides European-style dwellings housing just five or six people, fragmenting the large joint family groups traditional among the Nicobarese. "Who is helping whom?" Singh asks. "Are the victims helping the organizations to reproduce themselves?"

How can the messy problem of ineffective aid be fixed? Development analyst Ben Ramalingam has argued that we must understand development as a complex system to address underlying causes rather than treat symptoms. *Aftermath* shows Singh and colleagues in Vienna creating the Sustainable Indigenous Futures (SIF) fund in 2005, to provide financial aid directly to the community. As they soon learn, however, even well-directed intervention is no substitute for empowerment. Nominated community members struggle to abide by NGO norms of accountability. "A Nicobarese cannot become a project officer in ten days," Yusoof concedes. The community shuns Singh when the SIF stops sending money. In 2009, Singh and the SIF establish a partnership with a local NGO, the Tata Institute of Social Sciences, to reinstate resource-based livelihoods such as fishing in the islands.

Aftermath: The Second Flood
DIRECTOR: RAPHAEL BARTH
Golden Girls/Filmtank/Twopair/Tata Institute of Social Sciences/ORF: 2014.

Is the global aid enterprise learning from its mistakes? Perhaps. Extreme events can be catalysts for change. One young Nicobarese man leaves to pursue a higher degree, noting how the tsunami has shaped his aspirations. Yusoof is building a community tourist resort to generate income locally. Through such proactivity, resilience is built.

Finally, we see Singh, meticulously sorting slides, promising to document it all for the Nicobarese. But he questions the current paradigm of aid based on Western capitalist values. He even asks whether people in crisis need assistance at all. Scientists, he proposes, can help most by bringing together scientific and local ways of thinking to guide NGOs' actions on the ground.

Aftermath is a gritty, honest picture of two communities: the Nicobarese and aid agencies. Raw moments such as a ceremonial pig butchering, or a glimpse of the ethical and administrative conflicts that can trap development agencies in operational gridlock, are delivered unflinchingly. Ultimately, the Nicobarese have hope. "The aid has stopped," Yusoof concludes. "Now the real normal life starts, and we are happy." ■

Erin Bohensky *is a senior research scientist in livelihoods and adaptive development at the Commonwealth Scientific and Industrial Research Organisation in Townsville, Australia.*
e-mail: erin.bohensky@csiro.au

Correspondence

Target for ecosystem repair is impractical

The Convention on Biological Diversity (CBD) agreed in 2010 to restore "at least 15 per cent of degraded ecosystems" by 2020 (www.cbd.int/sp/targets). In Finland's experience, this target is unrealistic.

Ecosystem destruction is measured according to the expanse and degree of degradation. Consequently, under the CBD agreement, damage to an ecosystem's condition needs to be reduced by 15% over an entire area of degraded landscape or, for example, by 33% in a randomly selected 45% of that area to attain the same reduction.

There are considerable challenges in achieving this reduction: heavy restoration measures must be completed across large areas and in a short time, while compensating for ongoing degradation elsewhere.

Finland's forests cover roughly 15 million hectares, 95% of which have been degraded by forestry operations. To meet the CBD target, the country would need to reduce degradation by 33% in more than 100,000 hectares each month until 2020.

It has already taken Finland almost 30 years to restore some 30,000 hectares of forest (see Table 2.5 at go.nature.com/pajic9) — one-thousandth of the CBD's monthly requirement.

There is a danger that parties to the CBD will ignore the target because of its sheer impracticality.

Janne S. Kotiaho* *University of Jyväskylä, Finland.*
janne.kotiaho@jyu.fi
**On behalf of 7 correspondents (see go.nature.com/dcyoft for full list).*

Sports doping vastly underestimated

Roger Pielke Jr suggests that doping prevalence can be estimated by drug testing of athletes (*Nature* **517**, 529; 2015). I contend that this method is flawed: as the autobiographies of some athletes attest, regular dopers have a track record of avoiding testing positive.

To estimate doping's true prevalence, two procedures that circumvent inherent weaknesses in simple counts of positive test results are useful.

First, Bayesian inference methods can be used to compare two distributions of biological parameters affected by doping: one distribution among sampled athletes and the other in a suitable reference population, allowing the number of manipulated samples to be estimated (P.-E. Sottas *et al. Clin. Chem.* **57,** 762–769; 2011). A major advantage of this population-level analysis is that it can recognize abnormalities even when dopers' values remain within the 'normal' range.

Second, specially tailored questionnaires allow athletes to give honest answers to doping questions under cover of anonymity (W. Pitsch *et al. Eur. J. Sport. Soc.* **4,** 89–102; 2007). The questionnaires are based on the randomized response technique and are used by social scientists to study illegal and deviant behaviours.

A combination of these approaches estimates that 14–39% of elite athletes have intentionally doped (see O. de Hon *et al. Sports Med.* **45,** 57–69; 2015). This contrasts markedly with the 2% of samples designated as suspect in the World Anti-Doping Agency's published statistics.

Simon Evans *Uppsala University, Uppsala, Sweden.*
simon.evans@ebc.uu.se

A marine biologist's remarkable legacy

We write to dispel some myths in characterizations of the marine biologist Ed Ricketts, who died in 1948 (see, for example, A. Hirsh *Nature* **516,** 326–328; 2014).

Far from working in isolation, Ricketts interacted and corresponded with Torsten Gislén, George MacGinitie and Willis Hewatt, who all studied intertidal creatures. He did the same with scientists at the Smithsonian Institution in Washington DC and those at US and foreign universities. Never having taken a university degree, he was eager to tap into their expertise and to contribute to the international body of knowledge.

Hirsh's observation that Ricketts' book, *Between Pacific Tides,* was revolutionary because "it categorizes animals according to habitat, not phylum or family" is not the whole story. Ricketts' years of data from tidepool observations and collections resulted in a complex card-indexing and cross-referencing system that allowed him to build conclusions about species and their habitats and animal communities. As a result, his ecological data sets are still among the most robust ever to have been assembled.

Steven Albert, Mary Albert *Pacific Grove, California, USA.*
Don Kohrs *Hopkins Marine Station of Stanford University, Pacific Grove, California, USA.*
alberts1234@comcast.net

German initiative opens up animal data

Reliable and transparent data are essential to discussions of suffering in animal experimentation. A new initiative in Germany provides user-friendly public information about authorized animal-research projects.

The European directive to protect laboratory animals (2010/63/EU) requires researchers to provide an anonymous, non-technical summary of a proposed project, stating its purpose and potential benefits. The summaries also detail the number and types of animal to be used, the predicted harm to the animals and the evidence of compliance with the '3R' principles (see go.nature.com/jg7n6v). The severity of animal suffering and the likely human benefits are central to the approval process.

Germany's Federal Institute for Risk Assessment (BfR) has created a freely accessible, searchable website of these summaries to provide the public with clear insight into animal experiments (see www.animaltestinfo.de). Summaries are available to all EU member states, which could all set up similar open-access databases.

The summaries provide a unique channel for scientists to communicate their work to the public. They are a milestone in attempts to safeguard transparency in animal research, which is particularly controversial in the case of non-human primates.

Gilbert Schönfelder* *BfR, Berlin; and Charité — University Medicine Berlin, Germany.*
gilbert.schoenfelder@bfr.bund.de
**On behalf of 4 correspondents (see go.nature.com/ebzmxo for full list).*

Glass-blowing's Nobel moment

One of the most striking contributions of glass-blowing to research (see *Nature* **517,** 542–546; 2015) is the creation of 11,146 hemispherical photomultipliers in the neutrino detector used by Nobel laureate Masatoshi Koshiba. Their 50-centimetre diameters are a phenomenal achievement that called for great skill, strength and exceptional lung capacity.

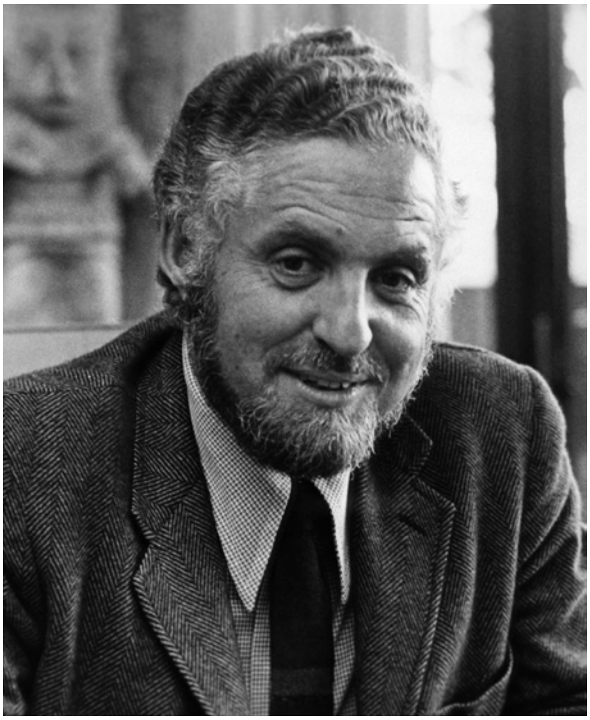
Koshiba was awarded the Nobel Prize in Physics in 2002 for the detection of cosmic neutrinos at the Kamioka Observatory in Hida, Japan. He acknowledged the glass-blowers of the Hamamatsu Photonics firm in his Nobel lecture, saying that the 50-cm-diameter photomultipliers "were the essential ingredient of the Kamioka experiments".

Min-Liang Wong *National Chung-Hsing University, Taichung, Taiwan.*
mlwong@dragon.nchu.edu.tw

Carl Djerassi

(1923–2015)

Chemist, writer and contraceptive-pill pioneer.



BETTMANN/CORBIS

Few scientists have changed society as much as Carl Djerassi did. By chemically synthesizing a steroid mimic of the hormone progesterone, Djerassi paved the way for the oral contraceptive pill, allowing women for the first time reliably to take control of their own reproductive choices. Djerassi's conviction that 'the pill' made the sexual liberalization of the 1960s possible is widely shared, and chemical control of the fertility cycle was a key ingredient in subsequent advances in reproductive technologies, beginning with *in vitro* fertilization (IVF) in the late 1960s.

This alone would have warranted Djerassi's acclaim as a scientist. But he also made important contributions to the synthesis of antihistamines and other natural products, pest control using hormone derivatives, mass spectrometry and computerized methods for the determination of complex molecular structures. He spoke and wrote extensively about the role of science in society and the ethical dilemmas it brings up. He also wrote fiction and plays. Calling himself an "intellectual polygamist", he showed how science could and should make its presence felt outside the laboratory.

Djerassi, who died on 30 January aged 91, was born to Jewish parents in Vienna in 1923. His parents separated in the late 1930s, and in 1939, following the Nazi annexation of Austria, he left with his mother for the United States. His PhD at the University of Wisconsin–Madison concerned the chemical transformation of human male and female sex hormones. In 1942, he joined the pharmaceutical company Ciba in New Jersey, where he took out a patent on one of the first synthetic antihistamines.

Djerassi started working at the small pharmaceutical company Syntex in Mexico City in 1949. There he established how to synthesize cortisone from a natural product derived from the Mexican yam. He then found that the same starting compound could yield norethisterone, a mimic of progesterone, which controls the female menstrual cycle. The discovery was developed in the late 1950s by biologist Gregory Pincus and gynaecologist John Rock, both of whom had conducted early work on IVF, to create a commercial contraceptive pill. Such pills were first approved in the United States in 1960.

The previous year, Djerassi had joined the chemistry department at Stanford University in California, where he stayed until he retired in 2002. His connection with the Syntex Corporation, of which he became director in 1960, and with the insect-control company Zoecon, which he founded in 1968, made him rich. Numerous awards followed.

Whereas many would have rested on their laurels, in his seventh decade Djerassi began a career as a writer and playwright, starting with the novel *Cantor's Dilemma* (1989). He styled himself an "intellectual smuggler", bringing science, under the guise of entertainment (which he called science-in-fiction), to audiences who might normally have recoiled from it.

His most successful play, *Oxygen* (2001), written with chemistry Nobel laureate Roald Hoffmann, explored the deliberations of a Nobel committee deciding whether to award a retrospective prize for the element's discovery to Antoine Lavoisier, Carl Wilhelm Scheele or Joseph Priestley. Like several of Djerassi's other works, it examined how scientific credit is awarded and the personal dynamics of research. The work of which he was most proud, however, was *Foreplay* (2011), which explored the sexual intrigues between a quartet of early-twentieth-century German intellectuals. His investment in the arts was profound and literal. Motivated by the suicide of his artist daughter Pamela in 1978, he founded an artists' colony in Woodside, California, which has hosted more than 2,000 residencies.

Even when Carl was forced to use a walking stick, it was easy to imagine that he had simply forbidden old age to plague him. "I had regaled my wife [Stanford professor of English Diane Middlebrook] for years with macho pride that I intended … to become the first non-retired centenarian professor at Stanford," he wrote. "I always allowed the barest of smiles to cross my face whenever I bragged in that fashion." That was Djerassi all over: proud to the point of arrogance, but at the same time mocking his own bombast.

No one who knew Carl would deny that there were prickly aspects to his character. He did not aspire to modesty and was quick to perceive slights. But his determination to grasp and grapple with what it means to be human in the modern age was admirable. His passion for the communication of scientific ideas was coupled with a profound aesthetic sensibility, as shown by his deep appreciation for the artist Paul Klee, several of whose works he owned.

Incredulity that Carl was never awarded a Nobel Prize was widely shared among his peers, and quite possibly by Carl himself. And his ambitions in writing and communication were not always welcomed even in his own institution, where he was once told that "in recent years your interests have been far removed from those most highly valued by your department".

Hopefully academia is more ready now to recognize the value, indeed necessity, of professional scientists who turn their energies to public engagement and to embedding science in its social context. Few have tried harder to do that than Carl Djerassi. ■

Philip Ball *is a science writer based in London.*
e-mail: p.ball@btinternet.com

For News & Views online, go to
nature.com/newsandviews

HIV

Tied down by its own receptor

An engineered protein that binds to the envelope of HIV viruses protects monkeys against infection with a simian–human virus that causes AIDS. This gene-therapy approach might provide an alternative to elusive HIV vaccines. SEE LETTER P.87

NANCY L. HAIGWOOD

The past 30 years have been marked by a long and discouraging search for an effective HIV vaccine. In 2009, the 'Thai trial' of the candidate vaccine RV144 was the first to demonstrate any success, measuring a 31.2% reduction in the rate of infection, although efficacy decreased over the first year after vaccination[1]. The difficulty in developing a more effective vaccine has forced investigators to explore problems that are posed by other intractable pathogens, including persistence in the host, a high degree of variability of certain regions, masking of common regions, and pathogen-induced inhibition of host immunity. But on page 87 of this issue, Gardner *et al.*[2] describe research suggesting that protection against HIV infection may be achievable through a gene-therapy approach, rather than by relying on eliciting protective immune responses by vaccination.

The trimeric envelope protein that is found on the surface of the viral particle of nearly all HIV strains binds directly to the CD4 receptor protein on the surface of many human immune cells, such as T cells and macrophages. This binding event causes a major shift in the envelope conformation, allowing the virus to bind to other co-receptors and enter the cell. It has been known since 1984 that CD4 is the receptor for HIV[3,4], and various forms of stabilized CD4 tethered to human immunoglobulin molecules (CD4–Ig) have been proposed and tested as potential therapeutics — the idea was that viral binding to these constructs would 'neutralize' the virus by preventing it binding to and entering cells. But this approach failed. Gardner and colleagues' findings provide the first logical explanation for this failure, and suggest an elegant way of using human CD4 derivatives to prevent infection.

The researchers engineered CD4 by fusing it with a mimetic of the amino terminus of CCR5, the host-cell co-receptor used by most HIV-1 strains during infection and disease progression. The CCR5 terminus has two sulfated tyrosine amino-acid residues that bind to the HIV envelope and facilitate viral entry[5], so the peptide mimetic is a sulfopeptide. The mimetic was based on an antibody that binds to the CCR5 binding site of the viral envelope; the authors modified and positioned it in the CD4–Ig construct for maximum activity and fit.

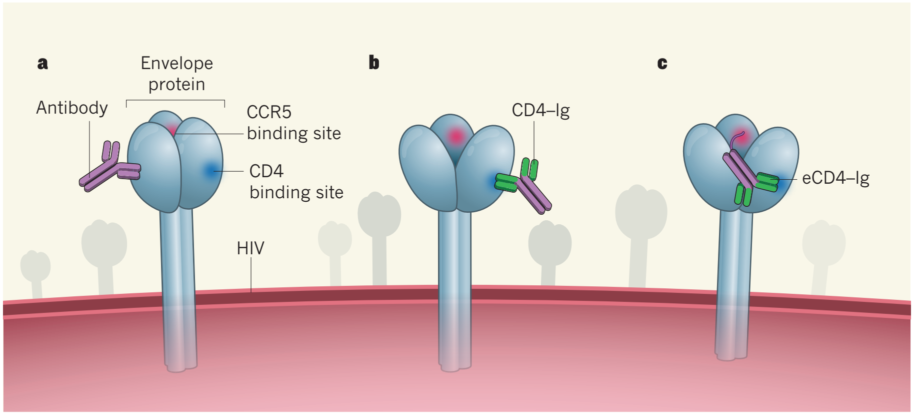


Figure 1 | Vaccination versus gene therapy. **a**, HIV infection begins with the virus's envelope protein binding to CD4 and CCR5 molecules on the surface of T cells. Most current strategies aimed at conferring protection against HIV focus on vaccines that are designed to prevent this binding by producing antibodies that bind to structures shared by the envelope of many HIV strains. However, antibodies bind only to small sections of the envelope, and the virus can evolve to shield these regions from antibody binding. **b**, An alternative approach proposed to stop virus binding to host cells is to use artificial constructs of human CD4 attached to immunoglobulin molecules. These CD4–Ig constructs will bind many viral strains, but they may expose the CCR5 binding site on the envelope protein and thus actually facilitate binding of the virus to CCR5 on the host cell. **c**, Gardner *et al.*[2] present an alternative construct, eCD4–Ig, which contains both CD4 and a mimetic of CCR5 and therefore blocks both points of viral binding.

This synthetic compound, named eCD4–Ig, has potent and broad neutralizing activity against all HIV isolates tested, including viral strains that are typically thought of as highly resistant to neutralization. It achieved these effects at lower concentrations than required when using the neutralizing monoclonal antibodies (NmAbs) that arise during the immune responses of some patients to HIV, and which are currently a major focus of attempts to develop HIV vaccines that prevent infection, rather than modulate viraemia once infection has occurred[6]. Furthermore, the construct was more effective than previous CD4–Ig constructs or the NmAb b12 at inducing immune killing of HIV-infected cells — a process known as antibody-dependent cellular cytotoxicity, which functions in concert with viral neutralization.

Gardner *et al.* went on to show that the eCD4–Ig construct imparted resistance to HIV-1 when infused into mice that model human HIV infections. As a further test of *in vivo* activity, the authors treated monkeys with an adenovirus-associated virus (AAV) that expressed the gene encoding a rhesus macaque version of the eCD4–Ig construct and with a separate AAV vector expressing a rhesus macaque enzyme to promote efficient sulfation. This gene-therapy vector allows continuous expression of the desired proteins in host cells by integrating into the host genome.

The animals expressed the transgene stably, although at different levels, and all were fully protected against repeated challenge with increasing doses of SHIV (a virus combining parts of the simian immunodeficiency virus (SIV) and HIV genomes). This protection was sustained for as long as 34 weeks after AAV transduction, and was achieved despite the monkeys receiving the virus intravenously, which is considered the infection route that provides the most stringent test of protection. These findings improve on an earlier test of the AAV transduction system to express a NmAb specific for SIV in monkeys[7], in which only a subset of monkeys that expressed the transgene were protected from SIV challenge.

Why did Gardner and colleagues' construct work? In a nutshell, it all comes down to the way that eCD4–Ig binds to the virus (Fig. 1). Human NmAbs that are able to neutralize a

broad range of HIV-1 strains do so by binding with very high affinity to shared viral structures (epitopes) that have precise but relatively small footprints. However, HIV has a variety of tricks to shield these shared epitopes from the immune system, although some infected individuals — referred to as elite neutralizers — do produce NmAbs of this sort. By contrast, CD4 binds to the envelope of all HIV-1 strains, albeit at lower affinity than these 'super potent' NmAbs. However, CD4 binding leads to a conformational change in the envelope that exposes the CCR5 binding site, thus potentially promoting HIV-1 infection in CCR5-expressing cells[8]. The modifications introduced by Gardner *et al.* into their eCD4–Ig construct seem to overcome this problem by preventing the engagement of envelope proteins with CCR5, while at the same time engaging multiple parts of the viral envelope, thereby increasing the binding power of their construct.

The study raises several questions and a few caveats. First, the modified protein is not natural and required the co-expression of an enzyme to perform the efficient addition of the sulfate moiety onto tyrosine residues. Second, the sample size of the monkey studies was quite small, and larger experiments in non-human primates are warranted. Furthermore, the intravenous challenge route, although rigorous, is not representative of the vast number of HIV-1 exposures worldwide, and it remains to be seen how expression of eCD4–Ig would affect virus challenges at mucosal sites, which better mimic natural routes of infection. It is also not yet clear whether the construct needs to be expressed close to the challenge sites. This, too, could be tested in non-human primate models.

Another major question rests in understanding the safety of eCD4–Ig in humans. Immune responses against the protein were elicited in the monkeys, albeit less strongly than against human NmAbs, and such responses could undermine its efficacy. But perhaps the greatest caveat to clinical application of the construct is how it, or future derivatives, will be used in humans. Such a complex molecule is unlikely to be administered repeatedly to those at risk of HIV infection, although that might be considered if it could be applied topically. The risks of expressing the construct as a transgene, in a similar manner to Gardner and colleagues' monkey experiments, are not known, and this approach would require careful and stepwise clinical safety testing. However, in the absence of a vaccine that can elicit broadly protective immunity and prevent infection, and given the lack of major breakthroughs on the horizon to provide one, the idea of conferring potent, sustained vaccine-like protection against HIV infection through gene therapy is certainly worth strong consideration. ■

Nancy L. Haigwood *is in the Division of Pathobiology & Immunology, Oregon National Primate Research Center, Oregon Health & Science University, Beaverton, Oregon 97006, USA.*
e-mail: haigwoon@ohsu.edu

1. Rerks-Ngarm, S. *et al. N. Engl. J. Med.* **361,** 2209–2220 (2009).
2. Gardner, M. R. *et al. Nature* **519,** 87–91 (2015).
3. Dalgleish, A. G. *et al. Nature* **312,** 763–767 (1984).
4. Klatzmann, D. *et al. Nature* **312,** 767–768 (1984).
5. Farzan, M. *et al. Cell* **96,** 667–676 (1999).
6. Hansen, S. G. *et al. Nature* **502,** 100–104 (2013).
7. Johnson, P. R. *et al. Nature Med.* **15,** 901–906 (2009).
8. Hoxie, J. A. *Annu. Rev. Med.* **61,** 135–152 (2010).

This article was published online on 18 February 2015.

MATERIALS SCIENCE

Nanoscale locomotion without fuel

Computer simulations have revealed a mechanism by which nanostructures of the material graphene can be driven in one direction by controlling the stiffness of the underlying substrate.

AMANDA S. BARNARD

The ability to move when and where we want is fundamental to our way of life, and our capacity for directing the natural motion of other objects and materials is essential for a range of technologies, from medicine to power generation. Although the same principles apply at the nanometre scale, miniature machines based on conventional macro-scale mechanisms have suffered from various problems, including lack of directional control, crippling frictional forces and permanent adhesion to adjacent components through strong chemical bonding. Writing in *Physical Review Letters*, Chang *et al.*[1] introduce a new way of moving nanoscale materials that overcomes some of these challenges, and that does not need an external power source to drive it.

Being able to control the motion of nanomaterials would be extremely useful for processes that require the delivery of molecules and other nanoscale objects, and for the functioning of nanodevices such as energy-conversion systems. With specific applications in mind, several techniques for moving various nanostructures have been proposed, using electrical currents[2] (or charge[3]), selective heating[4] or complicated chemical reactions[5]. But none of these methods is intrinsic — the nanostructures do not move spontaneously,

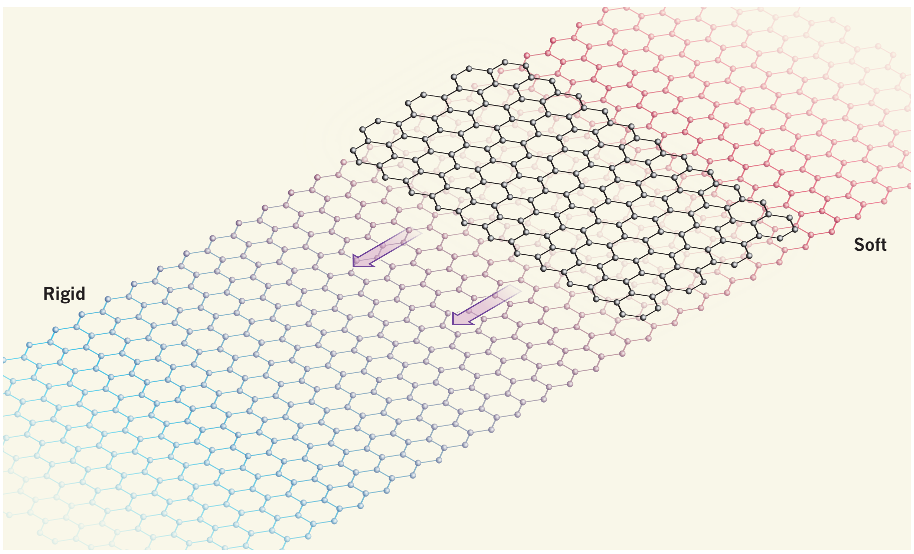


Figure 1 | Stiffness-guided motion. Chang and colleagues' computer simulations[1] reveal that, when a short graphene nano-flake (black) is placed on a graphene substrate containing a stiffness gradient, it spontaneously accelerates away from the soft (red) regions and towards more-rigid (blue) regions, without an external driving force. The authors attribute this behaviour to an inverse relationship between the substrate's stiffness and the interaction between the substrate and the flake (the van der Waals potential energy), which suggests that the velocity can be tuned. (Figure adapted from ref. 1.)

and each technique requires an energy source to sustain the motion. Furthermore, all the techniques can potentially damage the materials, which reduces the repeatability of each process. This is a problem, because viable nanotechnology must be reliable, reusable and cost-effective to run.

To find a method that fulfils these requirements and that has no need for external intervention, Chang *et al.* set up a series of virtual experiments using computer simulations. This approach offers some advantages over real-world experiments: one can be sure that the materials are free of defects and impurities, and that they are electrically, thermally and chemically isolated. Such isolation is particularly important when seeking intrinsic phenomena.

In their simulations, the authors first laid a nano-flake of graphene on top of a continuous graphene substrate, taking care to ensure that the lattices were not aligned. Graphene is a hexagonal lattice of carbon atoms, only one atomic layer thick, but its structure at the highly reactive edges varies depending on how the edges have been cut[6]. In the simulations, the flake was rotated by 30° with respect to the substrate, so that its edges did not line up with the lattice beneath. This reduced the risk of the flake becoming permanently attached to the substrate as a result of strong covalent bonding between the edges of the flake and the edges of the substrate.

Once it was in position, the authors made no further changes to the flake, but applied a stiffness gradient to the substrate, ranging from 0.801 to 4.005 newtons per metre along one direction. When the molecular-dynamics simulation was engaged, the nano-flake moved spontaneously — from a standing start — from the soft side to the hard side of the substrate (Fig. 1). Then, when it reached the end of the substrate, it rebounded because of a retraction force that pulled it back. Similar forces have been observed to pull extruded cores of multi-walled nanotubes back into the nanotubes[7].

Another advantage of computer simulations is that it allows animations of modelled processes to be made. In the present case, the overall motion is dramatically displayed in an animation provided with the paper's supplementary material. One can see that the graphene flake accelerates as it approaches the hard side of the substrate, and decelerates as it rebounds towards the soft side. This is clear evidence that the stiff side is energetically preferred.

Such stiffness-guided directional motion (termed durotaxis) was first observed in living cells, which also prefer rigidity[8]. Although the biological mechanism for durotaxis in cells remains a mystery, it has another similarity to the nanodurotaxis observed by Chang and co-workers in their simulated system: in both cases, weak van der Waals interactions are present, and in the latter case they were found to be crucial.

To prove this, the authors systematically repeated their virtual experiments under different simulation conditions, varying temperatures, stiffness gradients and stiffness configurations. The results unambiguously showed that the strength of the effective van der Waals potential — the interaction between the flake and the substrate — was inversely proportional to the stiffness. Lower potential energies are always more stable than higher ones, so this explains why the flake moves towards a rigid spot on the substrate: by doing so, it adopts a thermodynamically preferred state. At this stage, it is unclear whether perturbations to the system could be devised to reverse the motion, driving the flake back to soft regions.

Chang and colleagues' findings could have great potential in nanodevices, in part because the observed motion is conveniently unidirectional, but also because the underlying forces fall within a useful and technologically accessible range. The driving force for the nanoscale locomotion is about 320 kilopascals per square nanometre for a 6-nm-wide nanoflake on a stiffness gradient of 0.801–2.403 $N\,m^{-1}$. This is not too dissimilar from the forces in biological systems, such as the traction force per unit area exerted by a living cell on a substrate[8,9] and the driving force generated in a protein biomotor[10].

The challenge now is to fabricate graphene substrates that contain deliberate patterns of soft and hard regions, so that the experiments can be recreated in the real world. This will undoubtedly be difficult. It might also be possible to do this for other nanomaterials, but whether the simulated mechanism of nanodurotaxis will work for materials other than graphene is unknown. Nevertheless, the effort is certainly warranted, because strategic patterning of substrates might enable more-complicated trajectories to be realized, opening up new opportunities in nanoscale science and technology. ■

Amanda S. Barnard *is in the Virtual Nanoscience Laboratory, Commonwealth Scientific and Industrial Research Organisation, Parkville, Victoria 3052, Australia.*
e-mail: amanda.barnard@csiro.au

1. Chang, T., Zhang, H., Guo, Z., Guo, X. & Gao, H. *Phys. Rev. Lett.* **114,** 015504 (2015).
2. Dundas, D., McEniry, E. J. & Todorov, T. N. *Nature Nanotechnol.* **4,** 99–102 (2009).
3. Shklyaev, O. E., Mockensturm, E. & Crespi, V. H. *Phys. Rev. Lett.* **110,** 156803 (2013).
4. Barreiro, A. *et al. Science* **320,** 775–778 (2008).
5. van den Heuvel, M. G. L. & Dekker, C. *Science* **317,** 333–336 (2007).
6. Geim, A. K. & Grigorieva, I. V. *Nature* **499,** 419–425 (2013).
7. Zheng, Q. & Jiang, Q. *Phys. Rev. Lett.* **88,** 045503 (2002).
8. Lo, C.-M., Wang, H.-B., Dembo, M. & Wang, Y.-L. *Biophys. J.* **79,** 144–152 (2000).
9. Kim, J. H. *et al. Nature Mater.* **12,** 856–863 (2013).
10. Son, K., Guasto, J. S. & Stocker, R. *Nature Phys.* **9,** 494–498 (2013).

NEUROSCIENCE

A cellular basis for the munchies

How does marijuana cause the irresistible hunger pangs known as the munchies? Paradoxically, the answer seems to involve an unusual mode of activation of a brain circuit best known for suppressing appetite. **See Article p.45**

SACHIN PATEL & ROGER D. CONE

According to data gathered by the United Nations[1], 177 million people around the globe use marijuana. So some of you might be familiar with the munchies — that inexplicable drive to eat, stimulated by the active ingredients of marijuana, the cannabinoids. This connection has already led to the development of dronabinol, a synthetic version of the natural cannabinoid Δ-9-tetrahydrocannabinol, as a treatment for the metabolic disorder cachexia anorexia syndrome. But how, and where in the brain, do cannabinoids work to stimulate food intake? In this issue, Koch *et al.*[2] (page 45) report that, when given in doses meant to simulate the effects of marijuana, cannabinoids surprisingly activate a subset of pro-opiomelanocortin (POMC) neurons, a cell group in the brain's hypothalamus that has a central role in inhibiting hunger.

Previous work[3] has shown that ablation of POMC neurons, mutations in the gene encoding the POMC protein, and mutations in the melanocortin 4 receptor (MC4R) in downstream cellular targets of POMC neurons, all cause severe overeating (hyperphagia) and obesity. Conversely, experimental stimulation of the POMC cell group produces a slow-onset inhibition of food intake[4,5]. Cannabinoids bind to receptors dubbed CB_1Rs (for cannabinoid

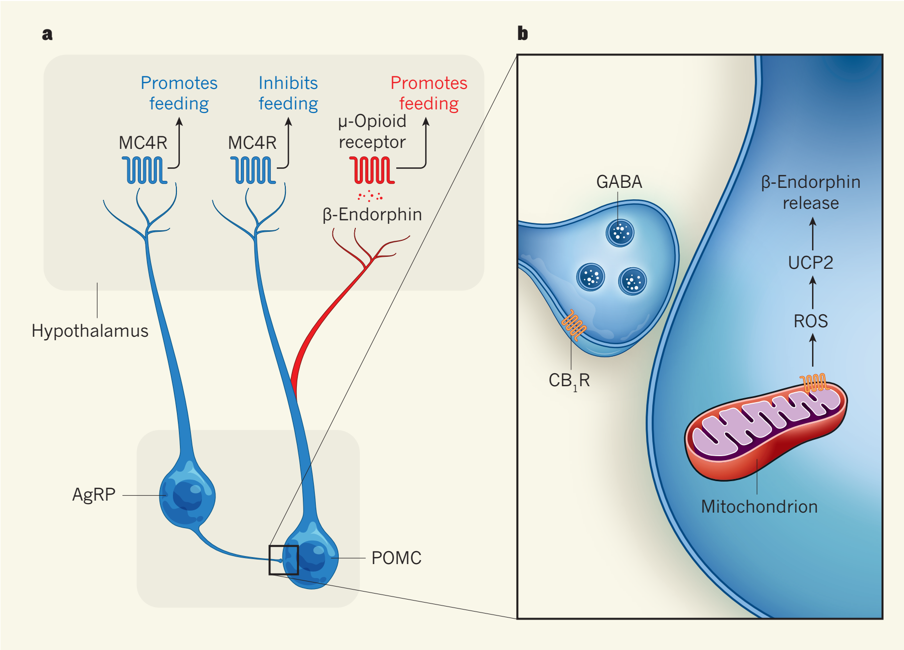


Figure 1 | Cannabinoid regulation of feeding circuits. **a**, The conventional view of appetite circuits in the brain's hypothalamus involves Agouti-related protein (AgRP) neurons, which stimulate feeding, and pro-opiomelanocortin (POMC) neurons, which inhibit it. Both groups affect downstream neurons expressing melanocortin 4 receptors (MC4Rs). Koch *et al.*[2] propose an alternative mechanism, in which cannabinoids stimulate food intake by causing a subset of POMC neurons to specifically activate β-endorphin-releasing boutons (release sites, not shown), which target downstream neurons expressing μ-opioid receptors. **b**, The authors suggest that, by activating cannabinoid receptor 1 (CB_1R) at two sites, cannabinoids increase feeding behaviour. At one site, they could inhibit release of the neurotransmitter GABA from AgRP neurons onto POMC neurons, thereby enhancing the latter neurons' excitability. At the other site, CB_1R activation in mitochondria could increase respiration, production of reactive oxygen species (ROS) and expression of mitochondrial uncoupling protein 2 (UCP2), which then acts as a switch to cause β-endorphin release selectively onto downstream neurons.

receptor 1), and Koch and colleagues observe that, in mice, these receptors are found on nerve terminals that make synaptic connections not only to POMC neurons, but also on organelles called mitochondria, in POMC neurons themselves. This binding stimulates the specific release of an orexigenic (appetite-stimulating) neuropeptide called β-endorphin from the neurons, while somehow avoiding release of α-melanocyte-stimulating hormone (α-MSH), an appetite-suppressing peptide found in these same neurons (Fig. 1a).

That cannabinoids can act at several brain regions to stimulate food intake is well established[6–8]. So the striking lesson here is not so much the orexigenic effect of cannabinoids through yet another of the many brain circuits involved in feeding behaviour. It is rather that cannabinoids can subvert an appetite-inhibitory (anorexic) circuit to become orexigenic, which indicates that the POMC circuit may be even more complex than previously thought. Another challenging concept arising from the present work is that, during the acute orexigenic response, cannabinoids stimulate these neurons partially through intracellular CB_1Rs, rather than through the more usually observed actions of cannabinoids at nerve terminals to regulate the release of substances such as the inhibitory neurotransmitter GABA.

It is noteworthy that, at a neuroanatomical level, the POMC circuit is quite complex, with POMC neurons belonging to the arcuate cluster of cells sending axonal projections to more than 100 brain regions. To add further complexity, a subset of these neurons responds to the hormone insulin; a different subset is affected by the hormone leptin[9], and roughly half of the cells may undergo inhibitory autoregulation by expressing the receptor MC3R (ref. 10). Moreover, almost all POMC neurons produce not only β-endorphin but also α-MSH.

These neurons may therefore be differentially secreting neuropeptides and neurotransmitters to either suppress or stimulate appetite. Parenthetically, administration of analogues of γ-MSH, another peptide released by POMC neurons, also has orexigenic effects[11], suggesting that β-endorphin may not be the only product of POMC neurons that can stimulate food intake. Thus, the emerging picture of the arcuate POMC system is that of a circuit that can sense a wide array of signals and can then produce highly discriminatory responses through a differentiated set of circuits and molecular signalling mechanisms.

What is remarkable about Koch and colleagues' findings is that cannabinoids seem to stimulate β-endorphin release selectively from POMC neurons. Consistent with this, the authors demonstrate that some 34% of the synaptic boutons, or release sites, on POMC neurons selectively express either β-endorphin or α-MSH.

There is, however, a nagging question about the proposed role of β-endorphin as the main mediator of the orexigenic actions of administered cannabinoids in POMC neurons. A previous study[12] found that deletion of the portion of the POMC gene encoding β-endorphin produces hyperphagia and obesity rather than leanness, as might be expected if the primary role of the natural peptide is orexigenic. This issue can be readily addressed by testing whether there is a decrease in the orexigenic response to cannabinoid administration in mice carrying mutations in β-endorphin, or perhaps in mice lacking the μ-opioid receptor, the target of β-endorphin in downstream neurons.

The present data certainly make a clear case for the striking ability of cannabinoids to stimulate a small subset of arcuate POMC neurons and subsequently to increase food intake. Nonetheless, the precise mechanisms of CB_1R-induced activation of POMC neurons and selective β-endorphin release remain to be fully elucidated.

The authors propose two potential routes by which CB_1R activation could boost POMC-neuron activity to increase feeding behaviour (Fig. 1b). First, low doses of CB_1R stimulators could increase firing in a subset of POMC neurons, most likely by reducing the release of incoming GABA signals that would otherwise dampen the neurons' activity[13]. Thus, these stimulators could modify the balance of excitation and inhibition in this neuronal subset. However, a previous study demonstrated[14] that CB_1R expression on neurons secreting the excitatory signal glutamate, rather than GABA, is required for the hyperphagic response to cannabinoids. Furthermore, increases in the firing rate of POMC neurons as a class reduces food intake[4], and thus cannot explain the hyperphagic effects of cannabinoids. So, the possibility of an alternative mechanism to cannabinoid-induced synaptic activation of a specific subset of POMC neurons requires further investigation.

Koch *et al.* suggest that activation of mitochondrial CB_1R represents just such an alternative. CB_1R activation has been shown to block respiration in mitochondria by inhibiting the cAMP–PKA signalling pathway[15]. The authors now extend these findings to show that low levels of CB_1R activation in fact increase mitochondrial respiration in POMC neurons as well as in neurons of the brain's hippocampus. They propose a signalling pathway involving CB_1R-induced increases in mitochondrial respiration, contact between mitochondria and another cellular organelle called the endoplasmic reticulum, generation of reactive oxygen

species and subsequent increased expression of mitochondrial uncoupling protein 2 (UCP2) — a regulator of both mitochondrial respiration in the hypothalamus and feeding (Fig. 1b). They show that UCP2 is essential for cannabinoid effects on mitochondrial respiration, β-endorphin release in the hypothalamus and feeding responses.

More definitive support for this provocative proposed mechanism could be provided by demonstrating that cell-impermeable inhibitors of CB_1R do not block feeding induced by CB_1R activators. Still unknown are the relative dominance of various CB_1R sites in the central nervous system in the orexigenic action of administered cannabinoids, and the relative importance of cell-surface and mitochondrial CB_1Rs. Regardless of this, Koch *et al.* provide another striking example of the complexity of the POMC circuits, and a new cellular mechanism by which cannabinoids stimulate feeding behaviour. ■

Sachin Patel *and* **Roger D. Cone** *are at the Vanderbilt University Medical Center, Department of Molecular Physiology and Biophysics, Nashville, Tennessee 37232-0615, USA.*
e-mail: roger.cone@vanderbilt.edu

1. United Nations Office on Drugs and Crime. *World Drug Report 2014* (UN, 2014).
2. Koch, M. *et al. Nature* **519,** 45–50 (2015).
3. Cone, R. D. *Nature Neurosci.* **8,** 571–578 (2005).
4. Aponte, Y., Atasoy, D. & Sternson, S. M. *Nature Neurosci.* **14,** 351–355 (2011).
5. Zhan, C. *et al. J. Neurosci.* **33,** 3624–3632 (2013).
6. Jamshidi, N. & Taylor, D. A. *Br. J. Pharmacol.* **134,** 1151–1154 (2001).
7. Kirkham, T. C., Williams, C. M., Fezza, F. & Di Marzo, V. *Br. J. Pharmacol.* **136,** 550–557 (2002).
8. Soria-Gómez, E. *et al. Nature Neurosci.* **17,** 407–415 (2014).
9. Williams, K. W. *et al. J. Neurosci.* **30,** 2472–2479 (2010).
10. Bagnol, D. *et al. J. Neurosci.* **19,** RC26 (1999).
11. Marks, D. L., Hruby, V., Brookhart, G. & Cone, R. D. *Peptides* **27,** 259–264 (2006).
12. Appleyard, S. M. *et al. Endocrinology* **144,** 1753–1760 (2003).
13. Ohno-Shosaku, T. & Kano, M. *Curr. Opin. Neurobiol.* **29,** 1–8 (2014).
14. Bellocchio, L. *et al. Nature Neurosci.* **13,** 281–283 (2010).
15. Bénard, G. *et al. Nature Neurosci.* **15,** 558–564 (2012).

This article was published online on 18 February 2015.

MOLECULAR BIOLOGY

Signals across domains of life

Signal sequences on messenger RNA that initiate protein synthesis are not thought to be interchangeable between life's domains. The finding that a signal from an arthropod virus can function in bacteria questions this idea. **SEE LETTER P.110**

ERIC JAN

All domains of life, from prokaryotes (archaea and bacteria) to eukaryotes (organisms that include plants, animals and fungi) use the ribosome apparatus to synthesize proteins by translating genetic code carried by messenger RNAs. Although the general steps of protein synthesis are evolutionarily conserved, the way in which ribosomes are recruited to an mRNA molecule differs depending on the specific phylogenetic domain[1]. In this issue, Colussi *et al.*[2] (page 110) reveal the surprising finding that a eukaryotic ribosome-recruiting signal is functional in prokaryotic bacteria, thereby challenging the prevailing dogma that prokaryotic and eukaryotic ribosome recruitment are mutually exclusive.

In prokaryotes, ribosomes are generally recruited by a specific signal on the mRNA called the ribosome-binding site (RBS, also called the Shine–Dalgarno sequence). The RBS positions the ribosome over the AUG start codon, a sequence that initiates protein synthesis[1]. By contrast, eukaryotic mRNAs do not contain an RBS, but instead contain covalent modifications — the 5′ cap at the 5′ end, and a poly(A) tail at the 3′ end — that promote the recruitment of ribosomes and protect the mRNA from degradation. Eukaryotes also use at least 12 protein initiation factors (compared to three in prokaryotes) that help to recruit the small subunit of the ribosome to the 5′ cap and facilitate scanning of the mRNA by the ribosome to find the AUG codon.

Some eukaryotic viral RNAs contain alternative signals called internal ribosome entry sites[3] (IRESs) to recruit the ribosome. When host-protein synthesis is inactivated during virus infection, an IRES can bypass normal signals, such as the 5′ cap and some initiation factors, to enlist the ribosome for viral protein synthesis. The simplest known IRES lies within regions found between the genes of dicistroviruses — a family of RNA viruses that infect arthropods[4].

Dicistrovirus IRESs are about 200 nucleotides long and fold into a unique RNA structure. What makes them so remarkable is their ability to bind directly to the ribosome without the need for any initiation factors and to initiate protein synthesis at a non-standard start codon[5] (not AUG). Structural[6] and biochemical studies[7] have shown that part of the IRES structure binds to the conserved core of the ribosome by mimicking a transfer RNA, which normally delivers an amino acid to the growing protein chain. The dicistrovirus IRES has therefore evolved to mimic a component of the normal translational machinery, permitting the viral RNA to hijack the ribosome.

Colussi and colleagues tested whether the dicistrovirus IRES can function in bacteria. To do this, they constructed mRNA that encodes reporter proteins — in this case, luminescent proteins. They found that inclusion of the IRES into the mRNA promotes expression of the reporter proteins in the bacterium *Escherichia coli*. Impressively, the researchers generated a comprehensive set of mutations within the IRES to work out how it functions in bacteria. They observed that some mutations that disrupt the IRES RNA structure

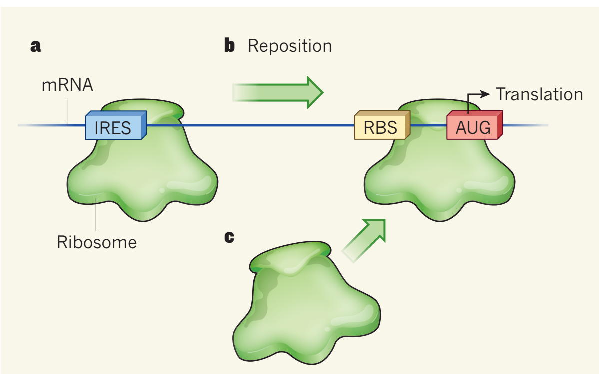


Figure 1 | Internal ribosome entry sites trigger protein synthesis in bacteria. Structures called internal ribosome entry sites (IRESs) in RNA viruses such as the dicistrovirus promote protein synthesis in eukaryotes (organisms including plants, animals and fungi). Colussi *et al.*[2] report the surprising finding that the dicistrovirus IRES can initiate protein synthesis in the bacterium *Escherichia coli* (a prokaryote). **a**, The authors constructed a messenger RNA that incorporates the IRES and introduced this to *E. coli* cells, where the IRES recruits the bacterial ribosome (the protein-synthesizing apparatus) to the mRNA. **b**, They propose that the ribosome then repositions itself to a ribosome-binding site (RBS) and an AUG start codon (an RNA sequence that initiates translation). **c**, Another hypothesis is that, after recruitment of the first ribosome, a second ribosome binds to the RBS and AUG codon.

reduced expression of the reporter proteins, demonstrating that the integrity of this structure is crucial for function.

To gain further insight into how IRESs that evolved to hijack ribosomes in eukaryotes can work in prokaryotic bacteria, Colussi *et al.* used X-ray crystallography to acquire a high-resolution image of the dicistrovirus IRES bound to the bacterial ribosome. Remarkably, the IRES binds to the ribosomal core, which is present across all domains of life, thus explaining how it can interact with both eukaryotic and bacterial ribosomes. The authors also show that an IRES structure from an unrelated virus — classical swine fever virus — does not support reporter-protein expression in bacteria, indicating that only the dicistrovirus IRES enables ribosome recruitment in *E. coli*.

But that is not the whole story. The investigators found that maximal reporter expression in bacteria depends not only on the IRES, but also on an RBS hidden in the reporter-system sequence and on an AUG start codon of the reporter mRNA. Moreover, their mutational analysis indicated that protein synthesis does not start at the IRES, but at the AUG codon. Colussi and co-workers therefore propose that, after being recruited to the IRES, the bacterial ribosome moves downstream to interact with the RBS and to start protein synthesis at the AUG codon (Fig. 1). Alternatively, another ribosome might be recruited to the RBS and initiate protein synthesis, a model that needs to be examined further.

It therefore seems that the authors' reporter system uses a hybrid of eukaryotic and prokaryotic signals to promote protein synthesis in bacteria. In prokaryotic translation, RNA structures are dynamic, and some can control protein expression by either burying the RBS within their structure or altering conformation to expose the RBS for ribosome recruitment[8]. Although leaderless RNAs — which are translated in the absence of the signals that usually support ribosome binding and translation efficiency — can also function across domains[9], Colussi *et al.* have shown for the first time that a bona fide signal in an RNA structure promotes protein synthesis in two domains of life.

This exciting study raises new questions. If the bacterial ribosome repositions itself from the IRES to the RBS, as the authors suggest, then how does it do so? Is the movement akin to ribosomal scanning of mRNAs in eukaryotes? And does the IRES function in bacteria without initiation factors, as it does in eukaryotic cells? Colussi and colleagues report that the dicistrovirus IRES binds to bacterial ribosomes slightly differently and more transiently than it does to eukaryotic ribosomes. Importantly, the IRES interacts with eukaryotic ribosomal protein rpS25, which is not present in bacterial ribosomes[10]. Further studies are needed to address how the IRES manipulates bacterial ribosomes.

Although the IRES functions in bacteria, does this phenomenon have a physiological role, and do other bacterial RNA structures function similarly? Another question is whether the IRES could work in archaea, or interact with the prokaryotic-like ribosome of mitochondria (specialized organelles that act as cellular powerhouses). And could the IRES be a remnant of a molecular fossil of the ancient 'RNA world' that is widely presumed to have preceded the evolution of DNA and proteins? Finally, this study opens up the possibility that other RNA structures function as signals across domains, and that eukaryotes and bacteria have more in common than was previously thought. ■

Eric Jan *is in the Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada.*
e-mail: ej@mail.ubc.ca

1. Malys, N. & McCarthy, J. E. G. *Cell. Mol. Life Sci.* **68,** 991–1003 (2011).
2. Colussi, T. M. *et al. Nature* **519,** 110–113 (2015).
3. Hellen, C. U. & Sarnow, P. *Genes Dev.* **15,** 1593–1612 (2001).
4. Bonning, B. C. & Miller, W. A. *Annu. Rev. Entomol.* **55,** 129–150 (2010).
5. Wilson, J. E., Pestover, C. V., Hellen, C. U. & Sarnow, P. *Cell* **102,** 511–520 (2000).
6. Fernández, I. S., Bai, X. C., Murshudov, G., Scheres, S. H. & Ramakrishnan, V. *Cell* **157,** 823–831 (2014).
7. Pfingsten, J. S., Costantino, D. A. & Kieft, J. S. *Science* **314,** 1450–1454 (2006).
8. Breaker, R. R. *Cold Spring Harb. Perspect. Biol.* **4,** 003566 (2012).
9. Moll, I., Grill, S., Gualerzi, C. O. & Bläsi, U. *Mol. Microbiol.* **43,** 239–246 (2002).
10. Nishiyama, T., Yamamoto, H., Uchiumi, T. & Nakashima, N. *Nucleic Acids Res.* **35,** 1514–1521 (2007).

This article was published online on 4 February 2015.

MATERIALS SCIENCE

Gating mechanism under pressure

Liquid-filled pores in membranes have been designed to reversibly open and close, allowing only particular fluids through at given pressures. This enables tunable and gated separations of mixtures of immiscible fluids. SEE LETTER P.70

MATHIAS ULBRICHT

Synthetic membranes, filters and microvalves have been developed for applications such as mass separation, analytics and medical therapies. But there is great interest in making these engineered systems 'smarter', by enabling biomimetic, tunable and gated selective permeation through a barrier. On page 70 of this issue, Hou *et al.*[1] report the proof of principle for such a gating mechanism, using pores that respond to pressure by reversibly switching between a liquid-filled 'closed' state and an 'open' state in which the liquid wets the pores' surface.

Many sophisticated and efficient barrier systems are found in nature that define, organize and protect cells, tissue and organisms. The most intriguing feature of such membranes is their ability to allow the highly selective movement of molecules or particles, or of bulk fluids (gases or liquids), through the barrier. Biological membranes are also dynamic: their permeability can respond to signals, enabling gating mechanisms, and they can have non-fouling or self-cleaning properties. Some of these systems have liquid-filled pores, and range from subnanometre-sized water-selective channels to micrometre-scale pores in lungs.

The opening of liquid-filled pores at specific gas pressures has long been used as a way of characterizing synthetic membranes, such as polymeric sterile filters that have typical barrier pore diameters of 0.5–0.1 micrometres; the pressure at which gas bubbles start to evolve from such membranes is related to the diameter of the largest pores or defects[2]. If the gas–liquid system in these tests is replaced by two immiscible liquids, smaller pores (down to a few nanometres) can also be analysed using pressures that will not mechanically damage the membrane, because the interfacial tension between two liquids is smaller than that between a gas and a liquid[3]. However, the pore-filling liquid is usually expelled during these tests, leaving the pore open for unrestricted fluid flow.

The reversibility and antifouling properties make these dynamic systems superior to conventional solid filter materials.

By contrast, Hou and colleagues' pore gating system (see Fig. 1b of the paper[1]) retains the pore-filling liquid, and is used in barriers that have pore sizes larger than 0.1 μm. The authors find that the pore-filling liquid must have a high affinity for the membrane polymer and be immiscible with the second liquid.

reduced expression of the reporter proteins, demonstrating that the integrity of this structure is crucial for function.

To gain further insight into how IRESs that evolved to hijack ribosomes in eukaryotes can work in prokaryotic bacteria, Colussi *et al.* used X-ray crystallography to acquire a high-resolution image of the dicistrovirus IRES bound to the bacterial ribosome. Remarkably, the IRES binds to the ribosomal core, which is present across all domains of life, thus explaining how it can interact with both eukaryotic and bacterial ribosomes. The authors also show that an IRES structure from an unrelated virus — classical swine fever virus — does not support reporter-protein expression in bacteria, indicating that only the dicistrovirus IRES enables ribosome recruitment in *E. coli*.

But that is not the whole story. The investigators found that maximal reporter expression in bacteria depends not only on the IRES, but also on an RBS hidden in the reporter-system sequence and on an AUG start codon of the reporter mRNA. Moreover, their mutational analysis indicated that protein synthesis does not start at the IRES, but at the AUG codon. Colussi and co-workers therefore propose that, after being recruited to the IRES, the bacterial ribosome moves downstream to interact with the RBS and to start protein synthesis at the AUG codon (Fig. 1). Alternatively, another ribosome might be recruited to the RBS and initiate protein synthesis, a model that needs to be examined further.

It therefore seems that the authors' reporter system uses a hybrid of eukaryotic and prokaryotic signals to promote protein synthesis in bacteria. In prokaryotic translation, RNA structures are dynamic, and some can control protein expression by either burying the RBS within their structure or altering conformation to expose the RBS for ribosome recruitment[8]. Although leaderless RNAs — which are translated in the absence of the signals that usually support ribosome binding and translation efficiency — can also function across domains[9], Colussi *et al.* have shown for the first time that a bona fide signal in an RNA structure promotes protein synthesis in two domains of life.

This exciting study raises new questions. If the bacterial ribosome repositions itself from the IRES to the RBS, as the authors suggest, then how does it do so? Is the movement akin to ribosomal scanning of mRNAs in eukaryotes? And does the IRES function in bacteria without initiation factors, as it does in eukaryotic cells? Colussi and colleagues report that the dicistrovirus IRES binds to bacterial ribosomes slightly differently and more transiently than it does to eukaryotic ribosomes. Importantly, the IRES interacts with eukaryotic ribosomal protein rpS25, which is not present in bacterial ribosomes[10]. Further studies are needed to address how the IRES manipulates bacterial ribosomes.

Although the IRES functions in bacteria, does this phenomenon have a physiological role, and do other bacterial RNA structures function similarly? Another question is whether the IRES could work in archaea, or interact with the prokaryotic-like ribosome of mitochondria (specialized organelles that act as cellular powerhouses). And could the IRES be a remnant of a molecular fossil of the ancient 'RNA world' that is widely presumed to have preceded the evolution of DNA and proteins? Finally, this study opens up the possibility that other RNA structures function as signals across domains, and that eukaryotes and bacteria have more in common than was previously thought. ■

Eric Jan *is in the Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada.*
e-mail: ej@mail.ubc.ca

1. Malys, N. & McCarthy, J. E. G. *Cell. Mol. Life Sci.* **68,** 991–1003 (2011).
2. Colussi, T. M. *et al. Nature* **519,** 110–113 (2015).
3. Hellen, C. U. & Sarnow, P. *Genes Dev.* **15,** 1593–1612 (2001).
4. Bonning, B. C. & Miller, W. A. *Annu. Rev. Entomol.* **55,** 129–150 (2010).
5. Wilson, J. E., Pestover, C. V., Hellen, C. U. & Sarnow, P. *Cell* **102,** 511–520 (2000).
6. Fernández, I. S., Bai, X. C., Murshudov, G., Scheres, S. H. & Ramakrishnan, V. *Cell* **157,** 823–831 (2014).
7. Pfingsten, J. S., Costantino, D. A. & Kieft, J. S. *Science* **314,** 1450–1454 (2006).
8. Breaker, R. R. *Cold Spring Harb. Perspect. Biol.* **4,** 003566 (2012).
9. Moll, I., Grill, S., Gualerzi, C. O. & Bläsi, U. *Mol. Microbiol.* **43,** 239–246 (2002).
10. Nishiyama, T., Yamamoto, H., Uchiumi, T. & Nakashima, N. *Nucleic Acids Res.* **35,** 1514–1521 (2007).

This article was published online on 4 February 2015.

MATERIALS SCIENCE

Gating mechanism under pressure

Liquid-filled pores in membranes have been designed to reversibly open and close, allowing only particular fluids through at given pressures. This enables tunable and gated separations of mixtures of immiscible fluids. See Letter p.70

MATHIAS ULBRICHT

Synthetic membranes, filters and microvalves have been developed for applications such as mass separation, analytics and medical therapies. But there is great interest in making these engineered systems 'smarter', by enabling biomimetic, tunable and gated selective permeation through a barrier. On page 70 of this issue, Hou *et al.*[1] report the proof of principle for such a gating mechanism, using pores that respond to pressure by reversibly switching between a liquid-filled 'closed' state and an 'open' state in which the liquid wets the pores' surface.

Many sophisticated and efficient barrier systems are found in nature that define, organize and protect cells, tissue and organisms. The most intriguing feature of such membranes is their ability to allow the highly selective movement of molecules or particles, or of bulk fluids (gases or liquids), through the barrier. Biological membranes are also dynamic: their permeability can respond to signals, enabling gating mechanisms, and they can have non-fouling or self-cleaning properties. Some of these systems have liquid-filled pores, and range from subnanometre-sized water-selective channels to micrometre-scale pores in lungs.

The opening of liquid-filled pores at specific gas pressures has long been used as a way of characterizing synthetic membranes, such as polymeric sterile filters that have typical barrier pore diameters of 0.5–0.1 micrometres; the pressure at which gas bubbles start to evolve from such membranes is related to the diameter of the largest pores or defects[2]. If the gas–liquid system in these tests is replaced by two immiscible liquids, smaller pores (down to a few nanometres) can also be analysed using pressures that will not mechanically damage the membrane, because the interfacial tension between two liquids is smaller than that between a gas and a liquid[3]. However, the pore-filling liquid is usually expelled during these tests, leaving the pore open for unrestricted fluid flow.

"The reversibility and antifouling properties make these dynamic systems superior to conventional solid filter materials."

By contrast, Hou and colleagues' pore gating system (see Fig. 1b of the paper[1]) retains the pore-filling liquid, and is used in barriers that have pore sizes larger than 0.1 μm. The authors find that the pore-filling liquid must have a high affinity for the membrane polymer and be immiscible with the second liquid.

The gating mechanism is based on established physical principles, and depends largely on the interfacial tensions between the pore surface, the pore-filling liquid and the fluid passing through the membrane. The pressure needed to open the pores (the gating threshold) to allow different bulk fluids to pass through can be adjusted by changing the polymeric membrane material, the pore size and the pore-filling liquid.

Hou and colleagues observed that the gating threshold can be varied over a wide range. Impressively, they show that systems can be set up so that air dispersed in water preferentially passes through the barrier under tunable conditions: at low pressure, the membrane is closed; beyond the critical pressure of air, all bubbles are removed from an air–water stream; and only beyond the second critical pressure does water also flow through the membrane. The reverse selectivity for such 'phase sorting' was also demonstrated, in which a gate opens at a lower pressure for a liquid than for a gas.

The authors report that three-component gas–liquid–liquid mixtures can also be separated by successively increasing the pressure across the membrane (see Fig. 3 of the paper[1]), and that the gating system has robust reversibility. Furthermore, they demonstrate that little fouling of the liquid-lined pore surface occurs when various particle-loaded liquids, including blood, are passed through the system. The reversibility and antifouling properties derive from the use of a liquid in the selective barrier, and make these dynamic systems superior to conventional solid filter materials.

The data suggest a range of applications for the authors' gating system, but there are potential challenges ahead. For example, long-term gating stability will be guaranteed only if the two liquids are completely immiscible. And the more complex the composition of the feed, the higher the risk that the pore-filling liquid will be extracted or polluted, which would change the interfacial tension. Moreover, achieving completely reversible reconfiguration of the pore-filling liquid will become more complicated as the aspect ratio (the ratio of length to width) of a pore increases, because some of the liquid is likely to be expelled. Designing a reservoir for temporarily expelled liquid could be a straightforward solution to this problem, however.

There are upper and lower limits to the pore sizes that can be used for phase sorting. For pores larger than a few tens of micrometres, the gating thresholds will be too low and close together to use for phase sorting. Lower limits in the submicrometre range are set by the membrane's stability to high gating thresholds, although this problem might be solved by using ceramic membranes instead of polymeric ones. Furthermore, when the thickness of the pore-lining liquid film is close to the pore radius, irregularities in pore size or shape will prevent defined gating. Membranes with well-controlled pore morphology might be needed.

Separations of homogeneous fluid mixtures are not possible using the new gating system, but stimulus-responsive membranes for ultrafiltration (which requires pore sizes between 50 nm and 2 nm) can be realized in a conceptually analogous way. For example, membranes have been reported[4] in which pore-filling gels reversibly switch between a closed ultrafiltering and an open macrofiltering state. Membranes formed from self-assembled polymers have also been made in which the nanoscale pore size can be tuned by applying a preselected transmembrane pressure[5].

So how might Hou and colleagues' system be used? Previous work[6] from the same research group described the preparation of slippery surface-immobilized liquid films that have good antifouling properties. If such films were used on the surface of pores or microfluidic channels, then it might be possible to make highly robust gating systems that have exceptionally good non-fouling properties. The integration of smart, liquid-filled porous valves in microfluidic systems could also enable the delivery of two-phase liquid mixtures at compositions and droplet sizes far from thermodynamic equilibria. This would be attractive for several applications, such as the *in situ* formulation of inks in advanced printing technologies. ■

Mathias Ulbricht *is at the Lehrstuhl für Technische Chemie II, Department of Chemistry, and the Center for Nanointegration Duisburg-Essen (CENIDE), Universität Duisburg-Essen, 45117 Essen, Germany.*
e-mail: mathias.ulbricht@uni-due.de

1. Hou, X., Hu, Y., Grinthal, A., Khan, M. & Aizenberg, J. *Nature* **519,** 70–73 (2015).
2. Mulder, M. H. V. *Basic Principles of Membrane Technology* 2nd edn (Kluwer, 1996).
3. Germic, L. *et al. J. Membr. Sci.* **132,** 131–145 (1997).
4. Adrus, N. & Ulbricht, M. *J. Mater. Chem.* **22,** 3088–3098 (2012).
5. Tyagi, P. *et al. Angew. Chem. Int. Edn* **51,** 7166–7170 (2012).
6. Wong, T.-S. *et al. Nature* **477,** 443–447 (2011).

CATALYSIS

Dual catalysis at the flick of a switch

A combination of two catalysts — one of which is light-activated — has been used to promote new chemical reactivity, opening up fresh opportunities for the synthesis of structurally complex organic molecules. **See Letter p.74**

JAMES J. DEVERY III
& COREY R. J. STEPHENSON

Difficulties can arise when developing reactions for the production of pharmaceuticals, industrial chemicals and materials if the chemical properties of the reaction components enable undesired products to form. Suppression of side reactions can therefore be just as important a consideration as how to obtain the synthetic target. Catalysis allows chemists to circumvent side products by inducing otherwise unreactive compounds to become selectively reactive. Even more control over reaction outcomes can be gained if reactivity can be induced by the flick of a light switch. On page 74 of this issue, Cuthbertson and MacMillan[1] report their use of visible-light-mediated catalysis to induce atypical activity of reaction substrates, initiating a complex transformation of otherwise unreactive components.

Photosynthesis is the archetypal example of how light can facilitate chemical reactions. During the day, plants absorb visible light and direct its energy to generate the molecules that fuel their lives, through chemistry that involves the transfer of single electrons. This intricate process uses light-activated catalysts to induce the desired electron transfer. Water acts as the source of electrons that allows photosynthetic complexes to complete their catalytic cycle[2]. This light-promoted system enables reactivity that chemists have striven to replicate for decades[3].

A form of light-activated catalysis called photoredox catalysis has become established as a powerful tool for initiating organic transformations mediated by free radicals[4]. Typical photoredox catalysts are transition-metal complexes or organic dyes that facilitate reactions in the presence of visible light. Methods using these compounds are at the forefront of single-electron transfer (SET) in organic synthesis because they are high-yielding, can be performed at room temperature and require fewer toxic reagents than other techniques. The fact that such 'mild' conditions can be used means that photoredox reactions are compatible with a wide variety of compounds.

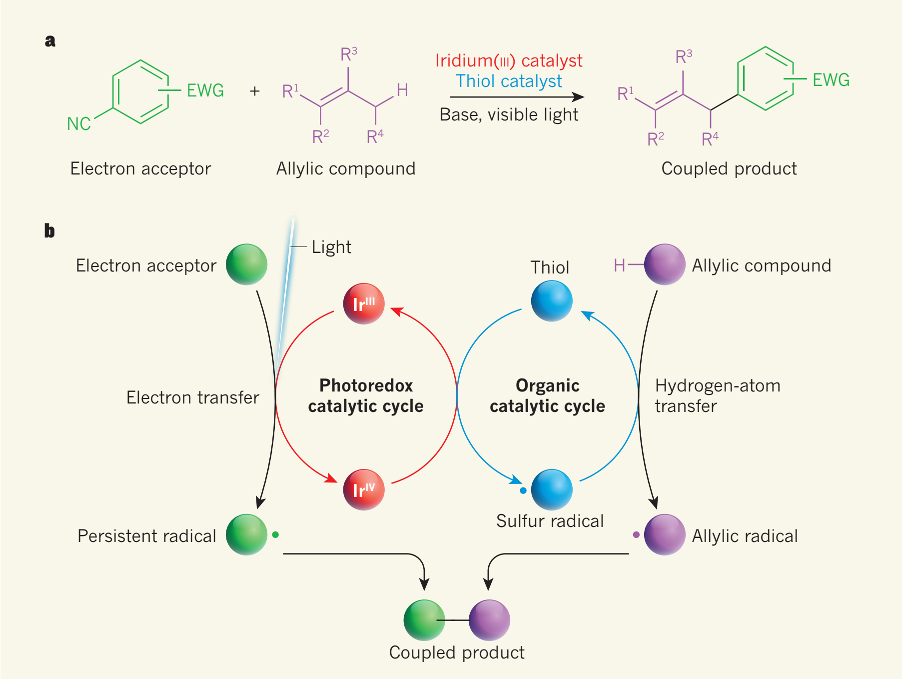


Figure 1 | Light-activated arylation of allylic carbon–hydrogen bonds. **a**, Cuthbertson and MacMillan[1] report that a combination of two catalysts — a visible-light-activated iridium(III) photoredox catalyst and an organic thiol catalyst — promote the reaction of electron-deficient aromatic compounds (which behave as electron acceptors) with allylic compounds in the presence of a base. Here, the aromatic compound contains a benzene ring; EWG represents an electron-withdrawing group, which can be attached at any position on the ring; R^1 to R^4 represent hydrocarbon groups. **b**, In this cartoon, the reaction components in **a** are represented as spheres for simplicity. When the iridium(III) catalyst is irradiated by light, the resulting excited complex transfers an electron to the electron acceptor, generating an iridium(IV) complex and a persistent radical. The iridium(IV) complex oxidizes the thiol to a sulfur radical, and is converted back to the original iridium(III) state. The sulfur radical removes a hydrogen atom from the allylic compound, re-forming the thiol and producing an allylic radical. Finally, the persistent radical and the allylic radical combine to form a coupled product.

Visible light is an advantageous energy source for reactions because it is inexpensive, is readily available and can be supplied by household light fixtures. Furthermore, visible-light-mediated SET processes can use inexpensive compounds as electron sources in a similar manner to the way plants use water. Finally, such reactions are particularly robust, with transformations requiring minimal amounts of catalyst.

Because photoredox catalysis is easy to use and beneficially directs the reactivity of molecules, it has allowed major advances to be made in organic synthesis when combined with other forms of catalysis[5–10]. Nevertheless, many of its applications are essentially alternative methods for performing well-established radical reactions. In particular, visible-light-mediated reactions have been used to replace radical processes that need expensive or toxic reagents[11]. But the true potential of photoredox catalysis lies in its ability to generate complex reactivity from simple substrates.

Cuthbertson and MacMillan realize this potential by combining a light-activated transition-metal catalyst with an organic catalyst to accomplish a reaction known as the arylation of allylic carbon–hydrogen bonds (Fig. 1a). They propose that this combination works through two linked catalytic cycles. Each cycle activates a different substrate, forming two radical intermediates, which combine to form the product.

The reaction builds on previously published work[12,13] from the same laboratory, and is triggered by a commercially available iridium catalyst. When illuminated, the catalyst donates an electron to a chemically reducible electron acceptor — specifically, an electron-deficient aromatic ring, such as a benzene ring with an electron-withdrawing group attached (Fig. 1b). This donation converts the electron acceptor into a radical species. The oxidized catalyst then gains an electron from a commercially available thiol (thiols are compounds containing SH groups), completing its catalytic cycle.

Meanwhile, the thiol forms a sulfur-containing radical that initiates the second catalytic cycle. The sulfur radical removes a hydrogen atom from an unsaturated hydrocarbon (an allyl-containing compound; allyl groups contain a CH–C=C motif); this step re-forms the thiol and generates an allylic radical. Finally, the radical resulting from the electron acceptor combines with the allylic radical to form a product. Thiols are generally used in radical chemistry as highly efficient and fast hydrogen-atom donors[14]. By contrast, the conditions used in Cuthbertson and MacMillan's system reverse this typical reactivity, inhibiting the thiol's ability to reattach a hydrogen atom to the allylic radical.

The power of this method lies in its use of readily available allylic hydrocarbon starting materials that are unreactive before exposure to light. Furthermore, it is compatible with relatively elaborate substrates, which react only at the positions activated by the catalysts — alternative reaction pathways that can often arise in complicated radical systems are avoided. Such selectivity is crucial if the reaction is to be used to synthesize intricate targets. Although the authors demonstrate a wide range of substrates, many opportunities for expansion of the reaction's scope remain.

The explosive growth of visible-light-mediated photoredox catalysis has yielded spectacular advances in methods for making chemical bonds that are difficult to access using conventional methods. Cuthbertson and MacMillan's work highlights the fact that new reactivity is also possible. However, there are still issues to be addressed with their method. Many substrates currently yield mixtures of isomeric products. Although this is ideal for generating libraries of molecules, it could potentially hamper applications in target-oriented synthesis. Further insight into the reaction mechanism of the linked catalytic cycles is also needed, and will provide the foundation for future development of this chemistry. ■

James J. Devery III *and* **Corey R. J. Stephenson** *are in the Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA.*
e-mail: crjsteph@umich.edu

1. Cuthbertson, J. D. & MacMillan, D. W. C. *Nature* **519,** 74–77 (2015).
2. Nelson, N. & Ben-Shem, A. *Nature Rev. Mol. Cell Biol.* **5,** 971–982 (2004).
3. Kärkäs, M. D., Verho, O., Johnston, E. V. & Åkermark, B. *Chem. Rev.* **114,** 11863–12001 (2014).
4. Prier, C. K., Rankic, D. A. & MacMillan, D. W. C. *Chem. Rev.* **113,** 5322–5363 (2013).
5. Nicewicz, D. A. & MacMillan, D. W. C. *Science* **322,** 77–80 (2008).
6. Kalyani, D., McMurtrey, K. B., Neufeldt, S. R. & Sanford, M. S. *J. Am. Chem. Soc.* **133,** 18566–18569 (2011).
7. DiRocco, D. A. & Rovis, T. *J. Am. Chem. Soc.* **134,** 8094–8097 (2012).
8. Hamilton, D. S. & Nicewicz, D. A. *J. Am. Chem. Soc.* **134,** 18577–18580 (2012).
9. Bergonzini, G., Schindler, C. S., Wallentin, C.-J., Jacobsen, E. N. & Stephenson, C. R. J. *Chem. Sci.* **5,** 112–116 (2014).
10. Du, J., Skubi, K. L., Schultz, D. M. & Yoon, T. P. *Science* **344,** 392–396 (2014).
11. Baguley, P. A. & Walton, J. C. *Angew. Chem. Int. Edn* **37,** 3072–3082 (1998).
12. Qvortrup, K., Rankic, D. A. & MacMillan, D. W. C. *J. Am. Chem. Soc.* **136,** 626–629 (2014).
13. Hager, D. & MacMillan, D. W. C. *J. Am. Chem. Soc.* **136,** 16986–16989 (2014).
14. Newcomb, M. *Encyclopedia of Radicals in Chemistry, Biology and Materials* (Wiley, 2012).

ARTICLE

doi:10.1038/nature14260

Hypothalamic POMC neurons promote cannabinoid-induced feeding

Marco Koch[1,2], Luis Varela[1], Jae Geun Kim[1]†, Jung Dae Kim[1,3], Francisco Hernández-Nuño[1], Stephanie E. Simonds[4], Carlos M. Castorena[5], Claudia R. Vianna[5], Joel K. Elmquist[5], Yury M. Morozov[6], Pasko Rakic[6,7], Ingo Bechmann[2], Michael A. Cowley[4], Klara Szigeti-Buck[1], Marcelo O. Dietrich[1,6], Xiao-Bing Gao[1], Sabrina Diano[1,3,6] & Tamas L. Horvath[1,3,6,7]

Hypothalamic pro-opiomelanocortin (POMC) neurons promote satiety. Cannabinoid receptor 1 (CB_1R) is critical for the central regulation of food intake. Here we test whether CB_1R-controlled feeding in sated mice is paralleled by decreased activity of POMC neurons. We show that chemical promotion of CB_1R activity increases feeding, and notably, CB_1R activation also promotes neuronal activity of POMC cells. This paradoxical increase in POMC activity was crucial for CB_1R-induced feeding, because designer-receptors-exclusively-activated-by-designer-drugs (DREADD)-mediated inhibition of POMC neurons diminishes, whereas DREADD-mediated activation of POMC neurons enhances CB_1R-driven feeding. The *Pomc* gene encodes both the anorexigenic peptide α-melanocyte-stimulating hormone, and the opioid peptide β-endorphin. CB_1R activation selectively increases β-endorphin but not α-melanocyte-stimulating hormone release in the hypothalamus, and systemic or hypothalamic administration of the opioid receptor antagonist naloxone blocks acute CB_1R-induced feeding. These processes involve mitochondrial adaptations that, when blocked, abolish CB_1R-induced cellular responses and feeding. Together, these results uncover a previously unsuspected role of POMC neurons in the promotion of feeding by cannabinoids.

Feeding behaviour is under the control of hypothalamic circuits[1]. In the hypothalamic arcuate nucleus (ARC), Agouti-related peptide (AgRP)-expressing neurons, when activated, promote food intake[2,3], whereas POMC-producing neurons promote satiety[4]. Homeostatic feeding regulation can be disrupted by exogenous substances, such as cannabinoids[5]. Activation of CB_1R can lead to robust feeding despite animals being sated[6]. However, the role of cannabinoids in the control of hypothalamic feeding circuits remains enigmatic[5–10]. In this study, we interrogated whether CB_1R-mediated feeding in a satiety state is associated with suppressed activity of POMC neurons, and if so, whether the altered activity of these neurons is important for CB_1R-induced feeding.

CB_1R drives activation of POMC neurons

We found that the selective CB_1R agonist arachidonyl-2′-chloroethylamide (ACEA) induced a known[11] bimodal feeding response in fed mice (Extended Data Fig. 1a–c). Notably, hyperphagic stimulation of CB_1R resulted in activation of POMC neurons, as assessed by expression of the proto-oncogene cFOS (Fig. 1A, B). *Ex vivo* electrophysiological recordings from ACEA-treated mice confirmed POMC neuronal activation (Fig. 1C). Next, we analysed slices acutely treated with ACEA. In the presence of tetrodotoxin (TTX), which blocks action potentials, ACEA failed to directly alter the membrane current of POMC neurons (Fig. 1D, a). We found that without TTX, low doses of ACEA (200 nM) induced depolarization of POMC neurons (Fig. 1D, b), as reported earlier[12], whereas high ACEA doses (1 μM) resulted in hyperpolarization of POMC cells (Fig. 1D, c). As an anatomical substrate of these effects, we detected CB_1R immunolabelling in both GABAergic and glutamatergic presynaptic terminals of POMC neurons (Fig. 1D, d). In line with these findings, hyperphagic doses of CB_1R agonists ACEA or WIN-55,212-2 mesylate (hereafter referred to as WIN) (1 mg kg^{-1} body weight) induced cFOS expression in POMC neurons (Fig. 1E). By contrast, the ACEA dose that did not affect food intake (5 mg kg^{-1} body weight) also did not induce cFOS expression in POMC neurons (Fig. 1E).

ACEA injected into the ARC resulted in a feeding response similar to that seen after its peripheral injection (Extended Data Fig. 1d). This local ACEA injection also activated POMC neurons as assessed by expression of phosphorylated CREB (pCREB(Ser 133)) and cFOS immunolabelling (Fig. 1F, G). Administration of the inverse CB_1R agonist rimonabant (RIMO) into the ARC blocked stimulation of food intake by peripheral ACEA application (Extended Data Fig. 1e–g), and, RIMO-mediated CB_1R blockade augmented the inactivation of POMC neurons in fasted mice as assessed by cFOS expression (Fig. 1H, I and Extended Data Fig. 1h).

POMC neurons drive feeding by cannabinoids

To determine whether the paradoxical POMC neuronal activation triggered by CB_1R is relevant for feeding, we injected inhibitory or stimulatory DREADD (AAV-hM4Di-mCherry or AAV-hM3Dq-mCherry[3]) into the ARC of *Pomc-Cre* mice (Extended Data Fig. 2a, b).

Clozapine-*N*-oxide (CNO)-driven activation of inhibitory DREADD reduced the numbers of cFOS-immunolabelled neurons in the ARC (Extended Data Fig. 2c), and blocked ACEA-induced activation of POMC neurons in fed mice (Extended Data Fig. 2d, e). CNO-mediated inhibition of POMC neurons in vehicle-treated mice enhanced feeding 8 h after CNO administration (Extended Data Fig. 2f). Conversely, DREADD-mediated activation of POMC neurons in vehicle-treated mice suppressed feeding 8 h after CNO application (Extended Data

[1]Program in Integrative Cell Signaling and Neurobiology of Metabolism, Section of Comparative Medicine, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [2]Institute of Anatomy, University of Leipzig, 04103 Leipzig, Germany. [3]Department of Obstetrics, Gynecology and Reproductive Sciences, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [4]Obesity & Diabetes Institute, Department of Physiology, Monash University, Clayton, Victoria 3800, Australia. [5]Division of Endocrinology & Metabolism, Department of Internal Medicine, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [6]Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [7]Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, Connecticut 06520, USA. †Present address: Division of Life Sciences, College of Life Sciences and Bioengineering, Incheon National University, Incheon 406-772, South Korea.

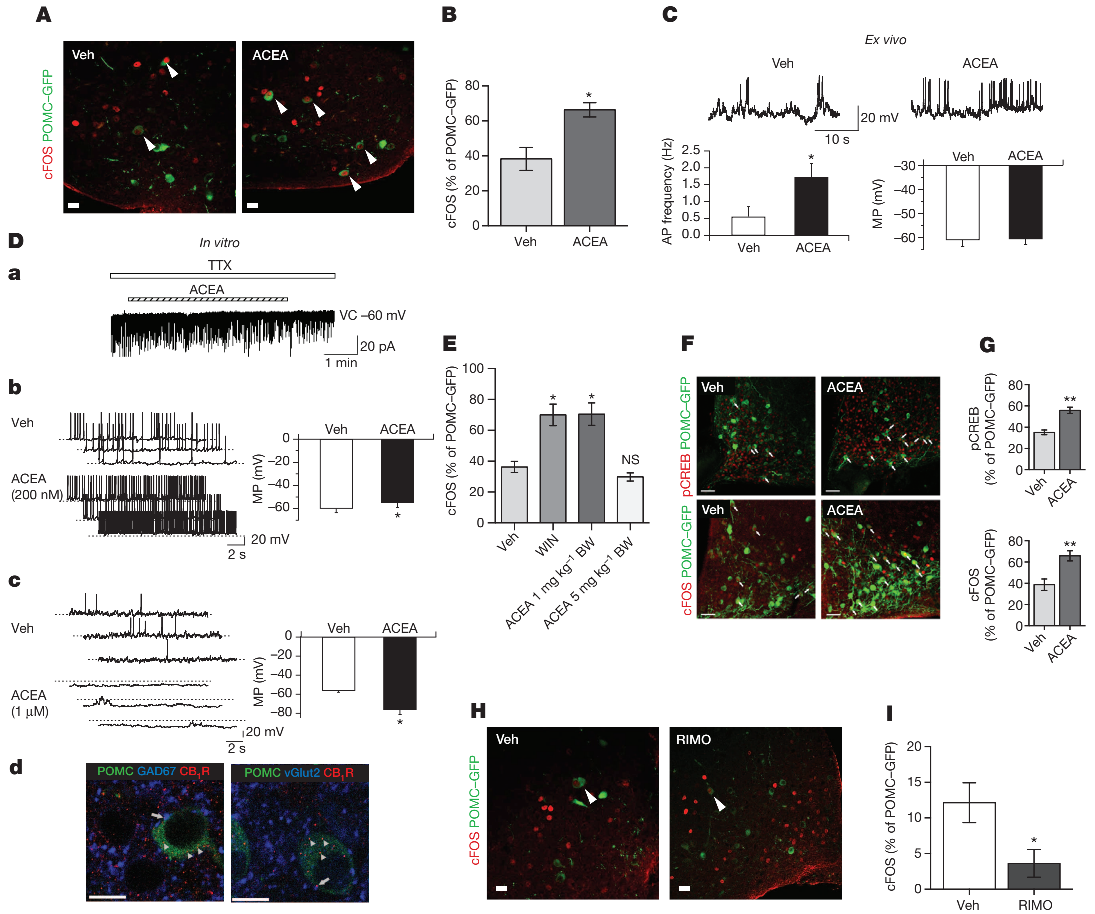


Figure 1 | CB$_1$R-driven paradoxical POMC activation. **A**, **B**, In fed mice, ACEA increased cFOS expression in POMC neurons (vehicle (Veh), $n = 5$ mice, 38.3 ± 6.6%; ACEA, $n = 4$, 66.4 ± 4.1%; $^*P < 0.05$). GFP, green fluorescent protein. **C**, In the fed state, ACEA increased POMC action potential (AP) frequency (left, vehicle, $n = 22$ cells per 7 mice, 0.56 ± 0.29 Hz; ACEA, $n = 22/7$, 1.73 ± 0.4 Hz; $^*P < 0.05$; seven independent experiments using litters from different parents). ACEA failed to alter POMC membrane potential (MP) (right, vehicle, −61.1 ± 2.7 mV; ACEA, −60.7 ± 2.3 mV). **D**, **a**, In the presence of TTX, 1 μM ACEA failed to alter POMC membrane current (4 out of 4 cells). VC, voltage clamp. **D**, **b**, Without TTX, 200 nM ACEA depolarized POMC neurons (vehicle, $n = 6$ cells per 3 mice, −59.8 ± 3.7 mV; ACEA, $n = 6/3$, −55.0 ± 4.2 mV; $^*P < 0.05$; three independent experiments). **D**, **c**, In the absence of TTX, a high ACEA dose (1 μM) hyperpolarized POMC neurons (vehicle, $n = 6$ cells per 4 mice, −56.2 ± 1.6 mV; ACEA, $n = 6/4$, −76.1 ± 5.1 mV; $^*P < 0.05$; four independent experiments). **D**, **d**, Arrows: CB$_1$R immunolabelling (red) in GABAergic (blue, left) and glutamatergic (blue, right) presynaptic terminals of POMC–GFP neurons (green). Arrowheads: POMC-GFP CB$_1$R immunolabelling (images are representative of three independent experiments). **E**, Increased cFOS expression in POMC neurons by hyperphagic but not neutral CB$_1$R activation (vehicle, $n = 3$ mice, 36.3 ± 3.6%; WIN, $n = 4$, 69.9 ± 7%; 1 mg kg^{-1} body weight (BW) ACEA, $n = 3$, 70.4 ± 7.2%; 5 mg kg^{-1} body weight ACEA, $n = 3$, 29.8 ± 2.6%; $^*P < 0.05$ versus vehicle (one-way analysis of variance (ANOVA) followed by Dunnett's multiple comparisons test). NS, not significant ($P > 0.05$). **F**, **G**, Local ARC hyperphagic CB$_1$R activation induced pCREB(Ser 133) (vehicle, $n = 4$ mice, 35.2 ± 2.2%; ACEA, $n = 5$, 55.8 ± 3%) and cFOS (vehicle, $n = 5$, 38.7 ± 5.4%; ACEA, $n = 5$, 65.8 ± 4.8%; $^{**}P < 0.01$) expression in POMC cells. **H**, **I**, In the fasted state, RIMO decreased POMC cFOS expression (vehicle, $n = 6$ mice, 12.1 ± 2.8%; RIMO, $n = 6$, 3.6 ± 1.9%; $^*P < 0.05$; two independent experiments). All values (biological replicates: **A**, **B**, **D**, **d**, **E–I**; biological replicates including technical replicates: **C**, **D**, **a–c**) denote mean ± s.e.m. POMC cFOS expression analyses in fed mice were performed in three independent experiments using litters from different parents (**A**, **B**, **E–G**). P values (unpaired comparisons) by two-tailed Student's t-test, unless otherwise stated. Scale bars, 25 μm (**A**, **D**, **d**, **H**) and 50 μm (**F**).

Fig. 2g). These findings are in line with previous observations that ARC POMC neurons gradually suppress feeding behaviour[4,13]. By contrast, DREADD-induced inhibition of POMC neurons significantly reduced acute, cannabinoid (WIN or ACEA)-induced feeding (Fig. 2a–d), whereas the acute hyperphagic response to the CB$_1$R agonist ACEA was significantly amplified by DREADD-induced activation of POMC neurons (Fig. 2e, f). Thus, our results demonstrate that POMC neuronal activation is key for feeding evoked by CB$_1$R in sated mice.

CB$_1$R drives β-endorphin but not α-MSH release

The *Pomc* gene encodes both the anorexigenic peptide α-melanocyte-stimulating hormone (α-MSH) and the opioid peptide β-endorphin, which can be orexigenic[14–16]. The hypothalamic paraventricular nucleus (PVN) is the main site where efferents of POMC neurons are thought to affect feeding[1,17,18]. ACEA-mediated activation of CB$_1$R did not affect α-MSH levels (Extended Data Fig. 3a–d, i and Extended Data Table 1a), but it was correlated with increased β-endorphin levels in the PVN[7] (Extended Data Fig. 3e–h, j and Extended Data Table 1b). In line with this, ACEA triggered the secretion of β-endorphin but not α-MSH into the supernatant of acute mediobasal hypothalamic explants derived from fed mice (Fig. 3a).

Next, we blocked CB$_1$R in the ARC by local injection of RIMO, which diminished ACEA-induced β-endorphin secretion (Fig. 3b). ACEA-induced β-endorphin secretion was absent in CB$_1$R knockout (*Cnr1*$^{-/-}$) mice (Fig. 3b). ACEA affected β-endorphin levels in a dose-dependent manner (Extended Data Fig. 4a and Extended Data Table 2), and local injection of ACEA into the ARC robustly induced β-endorphin levels in the PVN (Extended Data Fig. 4b, c and Extended Data Table 3).

α-MSH and β-endorphin are derived from POMC by pro-protein convertases, including PC-1 and PC-2 (ref. 14). CB$_1$R activation did

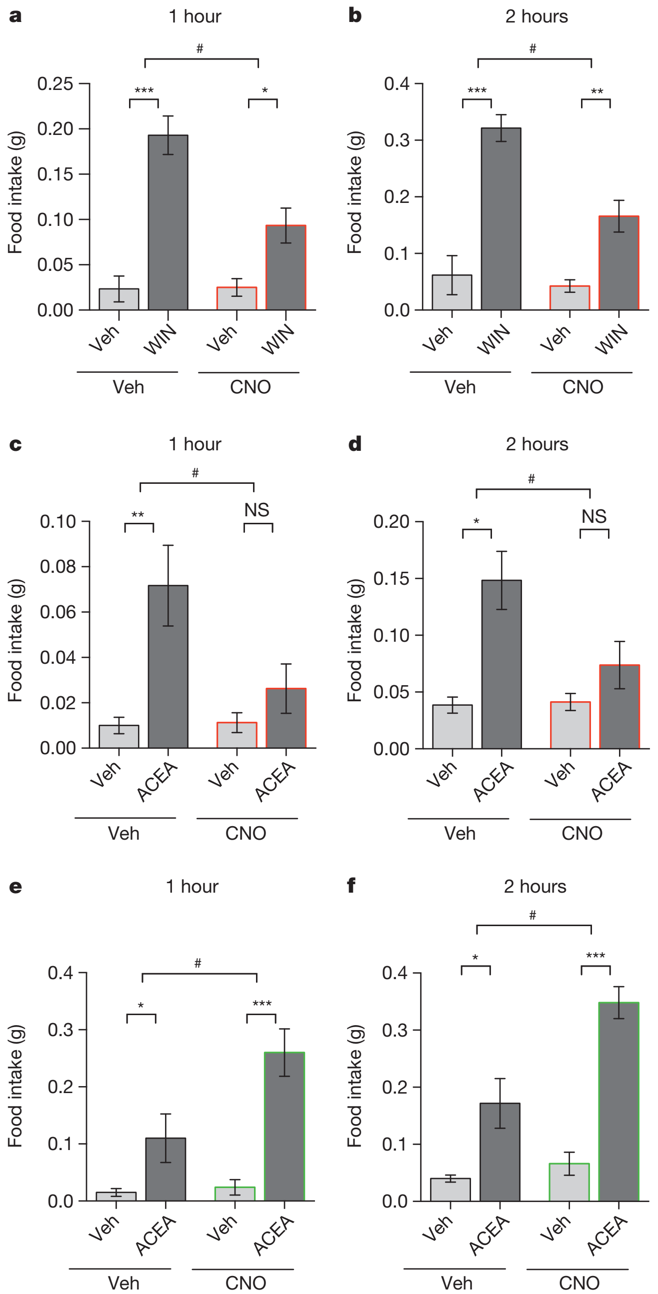


Figure 2 | DREADD-controlled POMC activity interferes with cannabinoid-induced feeding. **a**, **b**, DREADD-driven POMC inhibition reduced WIN-mediated hyperphagia, 1 h (**a**) and 2 h (**b**) after injection (**a**, 1 h: vehicle plus vehicle, $n = 6$ mice, 0.02 ± 0.01; vehicle plus WIN, $n = 13$, 0.19 ± 0.02; CNO plus vehicle, $n = 8$, 0.03 ± 0.01; CNO plus WIN, $n = 12$, 0.09 ± 0.02; **b**, 2 h: vehicle plus vehicle, 0.06 ± 0.03; vehicle plus WIN, 0.32 ± 0.02; CNO plus vehicle, 0.04 ± 0.01; CNO plus WIN, 0.17 ± 0.03; two independent experiments). **c**, **d**, DREADD-driven POMC inhibition blocked ACEA-mediated hyperphagia, 1 h (**c**) and 2 h (**d**) after injection (**c**, 1 h: vehicle plus vehicle, $n = 6$, 0.01 ± 0.004; vehicle plus ACEA, $n = 6$, 0.07 ± 0.02; CNO plus vehicle, $n = 8$, 0.01 ± 0.004; CNO plus ACEA, $n = 8$, 0.03 ± 0.01; **d**, 2 h: vehicle plus vehicle, 0.04 ± 0.01; vehicle plus ACEA, 0.15 ± 0.03; CNO plus vehicle, 0.04 ± 0.01; CNO plus ACEA, 0.07 ± 0.02; two independent experiments). **e**, **f**, DREADD-driven POMC activation enhanced ACEA-mediated hyperphagia, 1 h (**e**) and 2 h (**f**) after injection (**e**, 1 h: vehicle plus vehicle, $n = 6$, 0.02 ± 0.01; vehicle plus ACEA, $n = 6$, 0.11 ± 0.04; CNO plus vehicle, $n = 5$, 0.02 ± 0.01; CNO plus ACEA, $n = 5$, 0.26 ± 0.04; **f**, 2 h: vehicle plus vehicle, 0.04 ± 0.01; vehicle plus ACEA, 0.17 ± 0.04; CNO plus vehicle, 0.07 ± 0.02; CNO plus ACEA, 0.35 ± 0.03). All values (biological replicates) denote mean ± s.e.m. Five independent experiments using litters from different parents. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $\#P < 0.05$ interaction of WIN or ACEA and CNO (two-way ANOVA, followed by Šidák's multiple comparisons test).

not affect transcriptional regulation of either of these convertases (*Pcsk1* and *Pcsk2*, respectively; Extended Data Fig. 5a, b), but it did increase protein levels of PC-1 and PC-2 (Fig. 3c, d and Extended Data Fig. 5c).

Differential release dynamics of β-endorphin and α-MSH suggest that they may be present in non-overlapping intracellular compartments of POMC neurons. We observed that immunolabelling of these peptides did not overlap in $33.5 \pm 4.6\%$ (mean ± s.e.m.) of POMC boutons in the PVN (Fig. 3e). With electron microscopy, we confirmed that β-endorphin and α-MSH could be found in separate vesicles within the same neuronal profile of the PVN (Fig. 3f). Thus, our data indicate that separate vesicle pools of β-endorphin and α-MSH are under diverse regulatory mechanisms.

Naloxone blocks cannabinoid-induced feeding

Previous findings showed that local infusion of β-endorphin into the PVN could induce hyperphagia[19,20]. In fed mice, the μ-opioid receptor antagonist naloxone (NALO; 7.5 mg kg^{-1} body weight, intraperitoneal) diminished CB_1R-dependent hyperphagic response of WIN (Fig. 3g) and ACEA (Fig. 3h). Central application of naloxone into the PVN also diminished ACEA- and WIN-evoked hyperphagia (Fig. 3i).

β-endorphin was shown to be a feed-forward presynaptic inhibitor of AgRP neuronal activity[4]. In line with this, 90 min after ACEA application, the action potential frequency of AgRP neurons was decreased (Fig. 3j).

Taken together, we conclude that CB_1R-induced acute feeding in sated mice is evoked by POMC neurons via β-endorphin release and μ-opioid receptor activation while AgRP neurons are silenced.

CB_1R induces mitochondrial adaptations

Next we explored intracellular mechanisms that may bring about changes in neuropeptide release properties of POMC cells. Real-time PCR revealed local ARC CB_1R (*Cnr1*) expression (Fig. 4a). Immunolabelling for CB_1R showed both presynaptic and intracellular localization of CB_1R in POMC neurons (Fig. 4b, c). Electron microscopy revealed the presence of CB_1R immunoreactivity in mitochondria of POMC cells (Fig. 4d and Extended Data Fig. 5e), a finding consistent with previous observations of CB_1R in mitochondria of neurons[21–25]. Dynamic changes of mitochondria–endoplasmatic-reticulum (ER) contacts were shown in POMC neurons in association with changes in their activity and metabolic milieu[26,27]. In line with this, we found that hyperphagic activation of CB_1R resulted in an increased number of mitochondria–ER contacts (Fig. 4e–g). We also observed that ACEA-mediated activation of CB_1R dose-dependently increased hypothalamic coupled (state 3) and uncoupled (state 4) mitochondrial respiration (Fig. 4h–j). Increased coupled mitochondrial respiration is associated with increased generation of reactive oxygen species (ROS), which we showed previously to increase POMC neuronal activity[28]. Indeed, we found that hyperphagic ACEA administration increased ROS levels in POMC neurons (Fig. 5a, b).

CB_1R-induced energetic switch relies on UCP2

ROS are known to induce the expression and function of mitochondrial uncoupling protein 2 (UCP2)[29,30], a regulator of hypothalamic mitochondrial respiration and feeding[31,32]. Hyperphagic activation of CB_1R by ACEA increased hypothalamic expression of *Ucp2* messenger RNA (Fig. 5c). ACEA failed to alter mitochondrial respiration in the hypothalamus of UCP2 knockout mice (*Ucp2*$^{-/-}$; Fig. 5d). CB_1R-induced changes in POMC, PC-1 and PC-2 were diminished in *Ucp2*$^{-/-}$ mice (Fig. 5e–g and Extended Data Fig. 6c). Whereas CB_1R-driven POMC neuronal activation was partially retained (Fig. 5h), neither CB_1R-associated feeding (Fig. 5i) nor CB_1R-dependent increase of β-endorphin (Fig. 5j and Extended Data Table 4) was observable in *Ucp2*$^{-/-}$ mice.

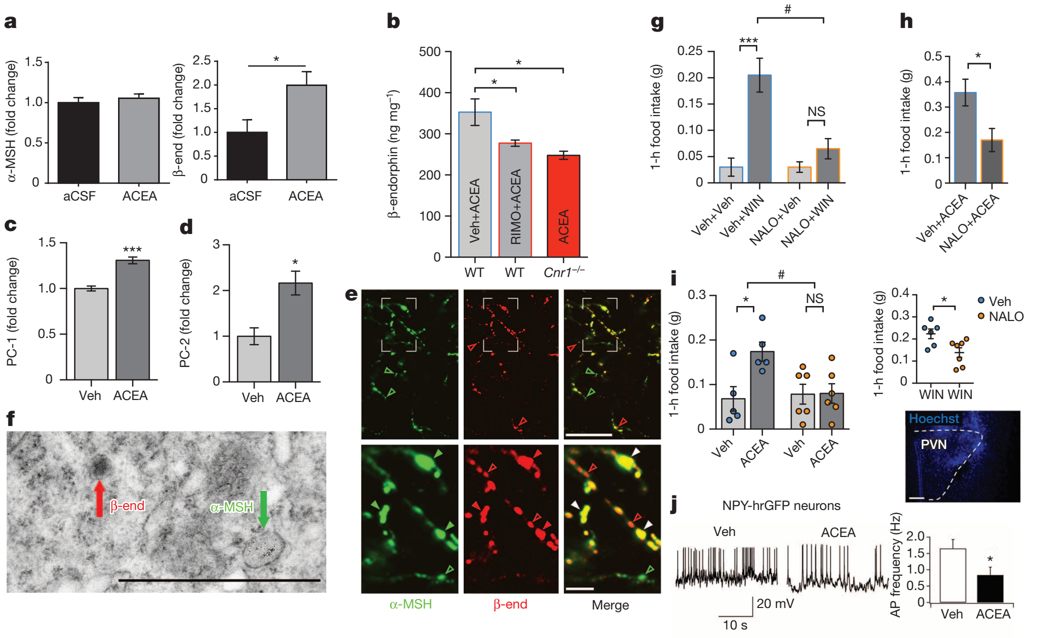


Figure 3 | CB_1R triggers hypothalamic β-endorphin release and drives feeding via opioid receptors. **a**, ACEA did not affect *in vitro* α-MSH secretion (artificial cerebrospinal fluid (aCSF), $n = 13$ mice, 1.00 ± 0.06; ACEA, $n = 6$, 1.05 ± 0.05; $P > 0.05$) but increased β-endorphin (β-end) release (aCSF, $n = 9$, 1.00 ± 0.27; ACEA, $n = 4$, 2.00 ± 0.29; $^*P < 0.05$). **b**, ARC administration of RIMO blocked ACEA-induced increase of β-endorphin as measured by ELISA in wild-type (WT) mice (vehicle plus ACEA, $n = 6$ mice, 352.7 ± 32.4; RIMO plus ACEA, $n = 6$, 277.4 ± 7.5; $^*P < 0.05$ versus vehicle plus ACEA). ACEA did not increase β-endorphin in CB_1R knockout (*Cnr1*$^{-/-}$) mice ($n = 4$, 247.8 ± 9.9; $^*P < 0.05$, one-way ANOVA, followed by Dunnett's multiple comparisons test). **c**, **d**, ACEA increased hypothalamic PC-1 (**c**) and PC-2 (**d**) protein levels (all groups $n = 4$ mice; PC-1: vehicle, 1.00 ± 0.03; ACEA, 1.31 ± 0.04; $^{***}P < 0.001$; PC-2: vehicle, 1.00 ± 0.18; ACEA, 2.16 ± 0.26; $^*P < 0.05$). **e**, Overview (top) and magnification (bottom) of α-MSH (green) and β-endorphin (red) in POMC fibres in the PVN. Coloured arrows indicate non-overlapping immunolabelling (images are representative of three independent experiments). **f**, Electron micrograph showing peroxidase immunolabelling of β-endorphin (red arrow) and immunogold labelling of α-MSH (green arrow) in different vesicles of the same process (images are representative of three independent experiments). **g**, **h**, Peripheral administration of NALO (7.5 mg kg^{-1} body weight, intraperitoneal) blocked hyperphagia induced by WIN (**g**; vehicle plus vehicle, $n = 10$ mice, 0.03 ± 0.02; vehicle plus WIN, $n = 10$, 0.21 ± 0.03; NALO plus vehicle, $n = 8$, 0.03 ± 0.01, NALO plus WIN, $n = 10$, 0.07 ± 0.02; $^{***}P < 0.001$, $\#P < 0.05$ for interaction of NALO and WIN; two-way ANOVA followed by Šidák's multiple comparisons test; two independent experiments) or induced by ACEA (**h**; vehicle plus ACEA, $n = 4$, 0.36 ± 0.05; NALO plus ACEA, $n = 4$, 0.17 ± 0.05; $^*P < 0.05$). **i**, PVN NALO application (5 μg per 0.5 μl) blocked hyperphagia induced by ACEA (intraperitoneal; vehicle plus vehicle, $n = 5$, 0.07 ± 0.03; vehicle plus ACEA, $n = 5$, 0.17 ± 0.02; NALO plus VEHI, $n = 6$, 0.08 ± 0.02; NALO plus ACEA, $n = 6$, 0.08 ± 0.02; $^*P < 0.05$, $\#P < 0.05$ for interaction of NALO and ACEA; two-way ANOVA, followed by Šidák's multiple comparisons test; two independent experiments) or induced by WIN (intraperitoneal; vehicle plus WIN, $n = 6$, 0.22 ± 0.02; NALO plus WIN, $n = 7$, 0.14 ± 0.02; $^*P < 0.05$; Hoechst (blue) injection for verification of correct cannula placement). **j**, Hyperphagic ACEA decreased AgRP/NPY-hrGFP (NPY-hrGFP) action potential (AP) frequency (vehicle, $n = 20$ cells per 6 mice, 1.65 ± 0.27 Hz; ACEA, $n = 20/6$, 0.84 ± 0.24 Hz; $^*P < 0.05$; six independent experiments using litters from different parents). All values (biological replicates: **a**–**i**; biological replicates including technical replicates: **j**) denote mean ± s.e.m. *P* values (unpaired comparisons) by two-tailed Student's *t*-test, unless stated otherwise. Scale bars, 25 μm (**e**, top), 5 μm (**e**, bottom), 1 μm (**f**) and 25 μm (**i**).

Thus, we conclude that UCP2 has a critical role in mediating CB_1R action on intracellular adaptation of POMC neurons while promoting feeding.

Discussion

Since the discovery of the melanocortin system and its satiety promoting action via melanocortin 4 receptors[33–35], POMC neurons have been considered as key drivers of the cessation of feeding[35–37]. However, our findings unmasked an overlooked, albeit previously proposed[16,38] role for hypothalamic POMC neurons in the promotion of feeding. Specifically, we found that POMC neuronal activation is indispensable for appropriate promotion of feeding triggered by CB_1R activation in the state of satiety. This observation brings these neurons to a new light in feeding regulation.

Cannabinoid-controlled feeding behaviour is complex and probably relies on diverse modes of action[11,39]. We focused on a specific action of cannabinoids, which is acute feeding evoked by CB_1R activation during the time of satiety. Thus, it is not unlikely that the cellular biological principles we uncovered here involving CB_1R action are not ubiquitous throughout the brain relating to cannabinoids. Nevertheless, the phenomenon of cannabis-triggered feeding in a state of satiety is a hallmark of marijuana use in humans[40,41], and whether hypothalamic POMC neurons may be relevant to other aspect of the complex behavioural responses triggered by cannabis use will be important to address.

We found that cannabinoid-induced activation of POMC neurons and related feeding relies on both pre-synaptic modalities and mitochondrial mechanisms. Regarding the presynaptic events, it is intriguing to note that CB_1R activation had a bimodal effect, in which lower doses of the CB_1R agonist ACEA induced depolarization of POMC neurons, while higher doses hyperpolarized these cells. Because both GABAergic and glutamatergic pre-synaptic terminals on POMC neurons expressed CB_1R, it is reasonable to assume that the bimodal effect of cannabinoids on feeding are due to the differential sensitivity of GABAergic versus glutamatergic axons to CB_1R activation[10,11].

The effect of CB_1R on mitochondria, and the mitochondrial localization of CB_1R are both in line with the original observations on CB_1R action in other brain sites[23]. In the present study, we showed that this CB_1R-dependent mitochondrial adaptation relies on the mitochondrial protein UCP2. UCP2 has been associated with hypothalamic feeding circuits[31,42–44], and was implicated in the malfunctioning of POMC neurons in diet-induced obesity[43]. However, most of those studies tied UCP2 activity to lipid and glucose use of hypothalamic neurons[32]. Here, we

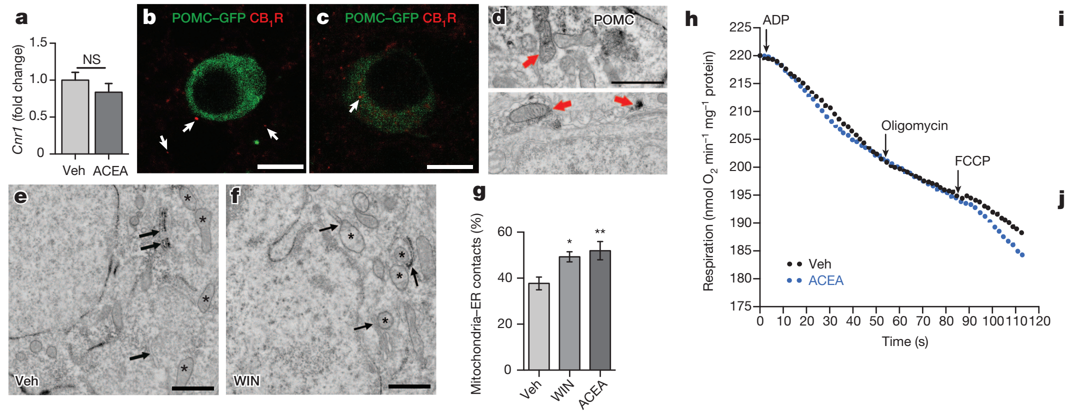


Figure 4 | CB_1R induces mitochondrial energetic switch in POMC neurons. **a**, ARC *Cnr1* mRNA expression (vehicle, $n = 5$ mice, 1.00 ± 0.11; ACEA, $n = 5$, 0.84 ± 0.12; $P > 0.05$). **b**, CB_1R (red; arrows) presynaptic to a POMC cell (GFP; green). **c**, **d**, Intracellular CB_1R in a POMC neuron (arrows). **d**, Mitochondrial CB_1R labelling (arrows) in POMC cell body (top) and process (bottom). Images are representative of three independent experiments. **e–g**, Increased mitochondria (stars)–ER (arrows) contacts in POMC neurons by WIN and ACEA (vehicle, $n = 10$ cells per 3 mice, 37.7 ± 2.8; WIN, $n = 10/3$, 49.3 ± 2.2; ACEA, $n = 10/3$ 51.9 ± 4; $*P < 0.05$, $**P < 0.01$ versus vehicle; one-way ANOVA, followed by Dunnett's multiple comparisons test). **h**, **i**, Hyperphagic ($1\ \mathrm{mg\,kg^{-1}}$ body weight) ACEA treatment increased *ex vivo* hypothalamic mitochondrial respiration (state 3: vehicle, $n = 6$ mice, 180 ± 8.9; ACEA, $n = 8$, 218.9 ± 15; state 4: vehicle, 81.2 ± 2.1; ACEA, 104.7 ± 5; $*P < 0.05$, $**P < 0.01$). **j**, Neutral dose of ACEA ($5\ \mathrm{mg\,kg^{-1}}$ body weight) reduced mitochondrial respiration (state 3: vehicle, $n = 8$, 186.1 ± 3.7; ACEA, $n = 6$, 156.8 ± 8.4; state 4: vehicle, 79.9 ± 4.7; ACEA, 63.8 ± 4.4). Three independent experiments using litters from different parents (**h–j**). All values (biological replicates: **a–d**, **h–j**; biological replicates including technical replicates: **e–g**) denote mean ± s.e.m. *P* values (unpaired comparisons) by two-tailed Student's *t*-test, unless stated otherwise. Scale bars, 25 μm (**b**, **c**), 0.5 μm (**d**) and 1 μm (**e**, **f**).

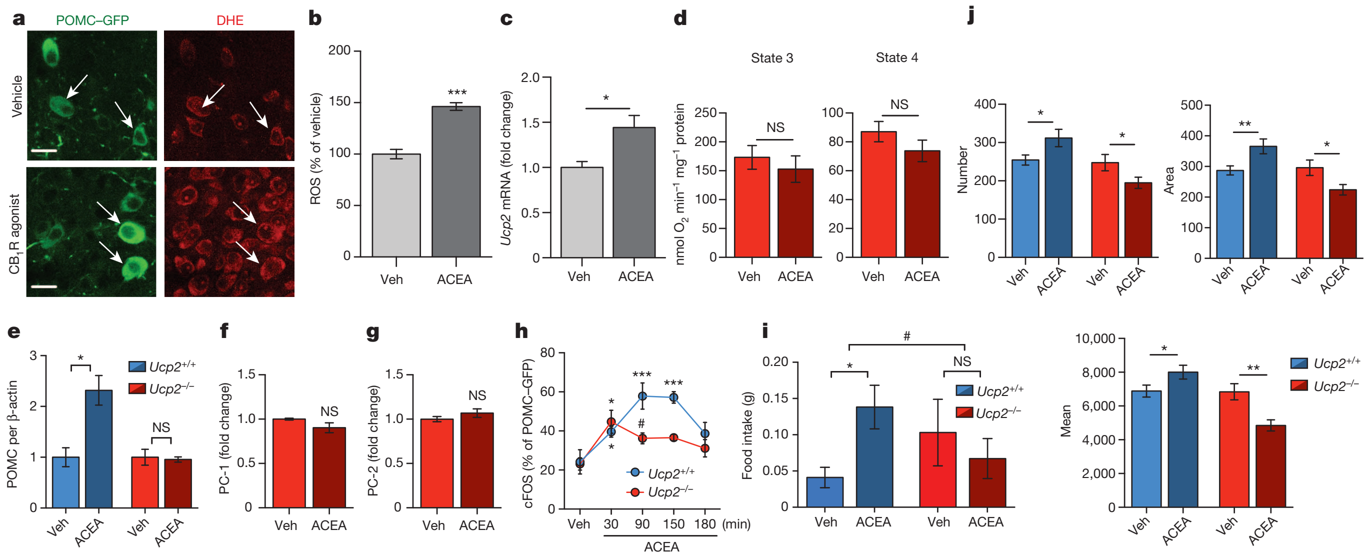


Figure 5 | CB_1R-induced energetic switch in POMC neurons relies on UCP2. **a**, **b**, ACEA increased POMC–GFP ROS (measured by dihydroethidium (DHE); vehicle, $n = 130$ cells per 7 mice, $100 \pm 4.5\%$; ACEA, $n = 248/9$, $146.2 \pm 3.7\%$; $***P < 0.001$; three independent experiments). **c**, ACEA triggered hypothalamic *Ucp2* mRNA expression (vehicle, $n = 5$ mice, 1.00 ± 0.07; ACEA, $n = 5$, 1.44 ± 0.13; $*P < 0.05$). **d**, No effect of ACEA on *ex vivo* mitochondrial respiration in $Ucp2^{-/-}$ mice (state 3: vehicle, $n = 3$ mice, 172.9 ± 20.6; ACEA, $n = 3$, 152.9 ± 22.9; state 4: vehicle, 87.1 ± 7.1; ACEA, 73.8 ± 7.5; $P > 0.05$). **e**, Increased POMC protein by ACEA in wild-type mice (vehicle, $n = 3$ mice, 1.00 ± 0.19; ACEA, $n = 3$, 2.3 ± 0.29; $*P < 0.05$) but not in $Ucp2^{-/-}$ littermates (vehicle, $n = 3$, 1.00 ± 0.16; ACEA, $n = 3$, 0.96 ± 0.05; $P > 0.05$). **f**, **g**, No effect of ACEA on PC-1 protein (**f**) (vehicle, $n = 3$ mice, 1.00 ± 0.01; ACEA, $n = 4$, 0.9 ± 0.06; $P > 0.05$) or PC-2 protein (**g**) (vehicle, $n = 3$, 1.00 ± 0.03; ACEA, $n = 4$, 1.07 ± 0.05; $P > 0.05$) in $Ucp2^{-/-}$ mice. **h**, cFOS expression in POMC neurons induced by ACEA in wild-type mice (vehicle, $n = 4$ mice, $22.1 \pm 6.2\%$; ACEA: 30 min, $n = 5$, $39.6 \pm 3\%$; 90 min, $n = 3$, $63.1 \pm 5{,}5\%$; 150 min, $n = 3$, $57.2 \pm 2.6\%$; 180 min, $n = 3$, $38.5 \pm 2.2\%$; $*P < 0.05$ versus vehicle, $***P < 0.001$ versus vehicle; one-way ANOVA, followed by Dunett's multiple comparisons test) and $Ucp2^{-/-}$ littermates (vehicle, $n = 6$ mice, $22.6 \pm 3.6\%$; ACEA: 30 min, $n = 6$, $42.2 \pm 6.4\%$; 90 min, $n = 6$, $36.6 \pm 3.1\%$; 150 min, $n = 3$, $37.3 \pm 1.6\%$; 180 min, $n = 3$, $33.5 \pm 5.2\%$; $*P < 0.05$ versus vehicle (one-way ANOVA, followed by Dunnett's multiple comparisons test); $\#P < 0.05$ versus wild type 90-min ACEA (unpaired comparisons by multiple *t*-tests, followed by correction for multiple comparisons using Holm–Šidák's method). **i**, ACEA (2 h)-induced hyperphagia in wild-type mice (vehicle, $n = 10$ mice, 0.04 ± 0.01 g; ACEA, $n = 11$, 0.14 ± 0.03 g; $*P < 0.05$) but not in $Ucp2^{-/-}$ littermates (vehicle, $n = 10$ mice, 0.1 ± 0.05 g; ACEA, $n = 14$, 0.07 ± 0.03 g; $P > 0.05$, $\#P < 0.05$ for interaction between ACEA and genotypes; three independent experiments using litters from different parents). All values (biological replicates: **c–i**; biological replicates including technical replicates: **a**, **b**) denote mean ± s.e.m. **j**, Increased PVN β-endorphin immunoreactivity in wild-type littermates (bilateral PVN analysis; vehicle, $n = 24$ values (technical replicates)/12 sections/4 mice (biological replicates); 90 min ACEA, $n = 24/12/4$); and reduced PVN β-endorphin immunoreactivity in $Ucp2^{-/-}$ mice (vehicle, $n = 18/9/3$; ACEA, $n = 18/9/3$; see Extended Data Table 4 for values; $*P < 0.05$, $**P < 0.01$). Values denote mean ± s.e.m.

identified UCP2 mediation as being crucial for the switch from α-MSH to β-endorphin release by POMC neurons triggered by CB_1R activation. This suggests a novel function of UCPs in the central nervous system in which they control the release of specific vesicle pools of neuropeptides in POMC neurons. Whether this effect of UCP2 is specific to POMC cells or represents a novel regulatory principle for other neuronal populations that express UCP2, including the midbrain dopamine system[45], needs to be explored. It is well known that a single neuron can have the capability of generating and releasing different types of neuropeptides[46]. How these different molecules with diverse functions are controlled within the same cells has not been fully resolved[46,47]. Our findings raise the possibility that changes in mitochondrial functions may be critical regulators of these processes.

In summary, the overall effect of cannabinoids on feeding seems to be driven by both pre- and post-synaptic effects, which may be independent from one another, and it is their temporal synchrony that brings about the overall behavioural changes.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 March 2014; accepted 23 January 2015.
Published online 18 February 2015.

1. Dietrich, M. O. & Horvath, T. L. Hypothalamic control of energy balance: insights into the role of synaptic plasticity. *Trends Neurosci.* **36,** 65–73 (2013).
2. Aponte, Y., Atasoy, D. & Sternson, S. M. AGRP neurons are sufficient to orchestrate feeding behavior rapidly and without training. *Nature Neurosci.* **14,** 351–355 (2011).
3. Krashes, M. J. *et al.* Rapid, reversible activation of AgRP neurons drives feeding behavior in mice. *J. Clin. Invest.* **121,** 1424–1428 (2011).
4. Yang, Y., Atasoy, D., Su, H. H. & Sternson, S. M. Hunger states switch a flip-flop memory circuit via a synaptic AMPK-dependent positive feedback loop. *Cell* **146,** 992–1003 (2011).
5. DiPatrizio, N. V. & Piomelli, D. The thrifty lipids: endocannabinoids and the neural control of energy conservation. *Trends Neurosci.* **35,** 403–411 (2012).
6. Bermudez-Silva, F. J., Cardinal, P. & Cota, D. The role of the endocannabinoid system in the neuroendocrine regulation of energy balance. *J. Psychopharmacol.* **26,** 114–124 (2012).
7. Sinnayah, P. *et al.* Feeding induced by cannabinoids is mediated independently of the melanocortin system. *PLoS ONE* **3,** e2202 (2008).
8. Bakkali-Kassemi, L. *et al.* Effects of cannabinoids on neuropeptide Y and β-endorphin expression in the rat hypothalamic arcuate nucleus. *Br. J. Nutr.* **105,** 654–660 (2011).
9. Ho, J., Cox, J. M. & Wagner, E. J. Cannabinoid-induced hyperphagia: correlation with inhibition of proopiomelanocortin neurons? *Physiol. Behav.* **92,** 507–519 (2007).
10. Hentges, S. T., Low, M. J. & Williams, J. T. Differential regulation of synaptic inputs by constitutively released endocannabinoids and exogenous cannabinoids. *J. Neurosci.* **25,** 9746–9751 (2005).
11. Bellocchio, L. *et al.* Bimodal control of stimulated food intake by the endocannabinoid system. *Nature Neurosci.* **13,** 281–283 (2010).
12. Hentges, S. T. Synaptic regulation of proopiomelanocortin neurons can occur distal to the arcuate nucleus. *J. Neurophysiol.* **97,** 3298–3304 (2007).
13. Zhan, C. *et al.* Acute and long-term suppression of feeding behavior by POMC neurons in the brainstem and hypothalamus, respectively. *J. Neurosci.* **33,** 3624–3632 (2013).
14. Dores, R. M. & Baron, A. J. Evolution of POMC: origin, phylogeny, posttranslational processing, and the melanocortins. *Ann. NY Acad. Sci.* **1220,** 34–48 (2011).
15. Dube, M. G., Horvath, T. L., Leranth, C., Kalra, P. S. & Kalra, S. P. Naloxone reduces the feeding evoked by intracerebroventricular galanin injection. *Physiol. Behav.* **56,** 811–813 (1994).
16. Kalra, S. P. & Horvath, T. L. Neuroendocrine interactions between galanin, opioids, and neuropeptide Y in the control of reproduction and appetite. *Ann. NY Acad. Sci.* **863,** 236–240 (1998).
17. Mountjoy, K. G., Mortrud, M. T., Low, M. J., Simerly, R. B. & Cone, R. D. Localization of the melanocortin-4 receptor (MC4-R) in neuroendocrine and autonomic control circuits in the brain. *Mol. Endocrinol.* **8,** 1298–1308 (1994).
18. Balthasar, N. *et al.* Divergence of melanocortin pathways in the control of food intake and energy expenditure. *Cell* **123,** 493–505 (2005).
19. Leibowitz, S. F. & Hor, L. Endorphinergic and alpha-noradrenergic systems in the paraventricular nucleus: effects on eating behavior. *Peptides* **3,** 421–428 (1982).
20. Leibowitz, S. F. Brain neurotransmitters and appetite regulation. *Psychopharmacol. Bull.* **21,** 412–418 (1985).
21. Morozov, Y. M. *et al.* Antibodies to cannabinoid type 1 receptor co-react with stomatin-like protein 2 in mouse brain mitochondria. *Eur. J. Neurosci.* **38,** 2341–2348 (2013).
22. Morozov, Y. M., Horvath, T. L. & Rakic, P. A tale of two methods: Identifying neuronal CB1 receptors. *Mol. Metabol.* **3,** 338 (2014).
23. Bénard, G. *et al.* Mitochondrial CB(1) receptors regulate neuronal energy metabolism. *Nature Neurosci.* **15,** 558–564 (2012).
24. Hebert-Chatelain, E. *et al.* Studying mitochondrial CB1 receptors: Yes we can. *Molecular Metabol.* **3,** 339 (2014).
25. Hebert-Chatelain, E. *et al.* Cannabinoid control of brain bioenergetics: Exploring the subcellular localization of the CB1 receptor. *Mol. Metabol.* **3,** 495–504 (2014).
26. Schneeberger, M. *et al.* Mitofusin 2 in POMC neurons connects ER stress with leptin resistance and energy imbalance. *Cell* **155,** 172–187 (2013).
27. Nasrallah, C. M. & Horvath, T. L. Mitochondrial dynamics in the central regulation of metabolism. *Nature Rev. Endocrinol.* **10,** 650–658 (2014).
28. Diano, S. *et al.* Peroxisome proliferation-associated control of reactive oxygen species sets melanocortin tone and feeding in diet-induced obesity. *Nature Med.* **17,** 1121–1127 (2011).
29. Negre-Salvayre, A. *et al.* A role for uncoupling protein-2 as a regulator of mitochondrial hydrogen peroxide generation. *FASEB* **11,** 809–815 (1997).
30. Echtay, K. S. *et al.* Superoxide activates mitochondrial uncoupling proteins. *Nature* **415,** 96–99 (2002).
31. Andrews, Z. B. *et al.* UCP2 mediates ghrelin's action on NPY/AgRP neurons by lowering free radicals. *Nature* **454,** 846–851 (2008).
32. Diano, S. & Horvath, T. L. Mitochondrial uncoupling protein 2 (UCP2) in glucose and lipid metabolism. *Trends Mol. Med.* **18,** 52–58 (2012).
33. Huszar, D. *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88,** 131–141 (1997).
34. Fan, W., Boston, B. A., Kesterson, R. A., Hruby, V. J. & Cone, R. D. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature* **385,** 165–168 (1997).
35. Cone, R. D. Anatomy and regulation of the central melanocortin system. *Nature Neurosci.* **8,** 571–578 (2005).
36. Elias, C. F. *et al.* Leptin differentially regulates NPY and POMC neurons projecting to the lateral hypothalamic area. *Neuron* **23,** 775–786 (1999).
37. Cowley, M. A. *et al.* Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature* **411,** 480–484 (2001).
38. Horvath, T. L., Naftolin, F., Kalra, S. P. & Leranth, C. Neuropeptide-Y innervation of beta-endorphin-containing cells in the rat mediobasal hypothalamus: a light and electron microscopic double immunostaining analysis. *Endocrinology* **131,** 2461–2467 (1992).
39. Soria-Gómez, E. *et al.* The endocannabinoid system controls food intake via olfactory processes. *Nature Neurosci.* **17,** 407–415 (2014).
40. Greenberg, I., Kuehnle, J., Mendelson, J. H. & Bernstein, J. G. Effects of marihuana use on body weight and caloric intake in humans. *Psychopharmacology* **49,** 79–84 (1976).
41. Foltin, R. W., Brady, J. V. & Fischman, M. W. Behavioral analysis of marijuana effects on food intake in humans. *Pharmacol. Biochem. Behav.* **25,** 577–582 (1986).
42. Coppola, A. *et al.* A central thermogenic-like mechanism in feeding regulation: an interplay between arcuate nucleus T3 and UCP2. *Cell Metab.* **5,** 21–33 (2007).
43. Parton, L. E. *et al.* Glucose sensing by POMC neurons regulates glucose homeostasis and is impaired in obesity. *Nature* **449,** 228–232 (2007).
44. Horvath, T. L. *et al.* Brain uncoupling protein 2: uncoupled neuronal mitochondria predict thermal synapses in homeostatic centers. *J. Neurosci.* **19,** 10417–10427 (1999).
45. Andrews, Z. B. *et al.* Ghrelin promotes and protects nigrostriatal dopamine function via a UCP2-dependent mitochondrial mechanism. *J. Neurosci.* **29,** 14057–14065 (2009).
46. Landry, M., Vila-Porcile, E., Hokfelt, T. & Calas, A. Differential routing of coexisting neuropeptides in vasopressin neurons. *Eur. J. Neurosci.* **17,** 579–589 (2003).
47. van den Pol, A. N. Neuropeptide transmission in brain circuits. *Neuron* **76,** 98–115 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank M. Shanabrough and J. Bober for technical support and R. Jakab for assisting with the illustrations. This work was supported by the US National Institutes of Health (DP1 DK098058, R01 DK097566, R01 AG040236 and P01 NS062686), the American Diabetes Association, The Klarmann Family Foundation, the Helmholtz Society (ICEMED) and the Deutsche Forschungsgemeinschaft SFB 1052/1 (Obesity Mechanisms).

Author Contributions M.K., S.D. and T.L.H. developed the conceptual framework of this study. M.K., M.O.D., X.-B.G., S.D. and T.L.H. interpreted results. M.K. performed experiments and analysed results. Experimental contributions: L.V. contributed to Figs 4h–j, 5d and Extended Data Figs 1b, 5e and 6a, b; J.G.K. contributed to Figs 2e, f, 3i, 5a, b and Extended Data Fig. 2g; J.D.K. contributed to Figs 3b–d, 5e–g and Extended Data Figs 5c and 6c; F.H. contributed to Figs 4a, 5c and Extended Data Fig. 5a, b, d; S.E.S. contributed to Fig. 3a; C.M.C., C.R.V. and J.K.E. provided key animal models; Y.M.M. and P.R. contributed to Fig. 3b and Extended Data Fig. 1c; P.R., I.B. and M.A.C. provided materials, animals and equipment; K.S.-B. contributed to Figs 3f and 4d–g; X.-B.G. contributed to Figs 1C, Da–c and 3j. M.K. and T.L.H. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.L.H. (tamas.horvath@yale.edu).

METHODS

All used mice were aged between 10 and 18 weeks at the time of killing. All procedures were approved by Yale University Institutional Animal Care and Use Committee (IACUC). Mice were maintained under standard laboratory conditions with water and food freely available unless otherwise stated. Mice were housed on a 12 h light/12 h dark cycle with lights on at 07:00 and off at 19:00. All experimental groups within one experiment contained mice of same strain and sex, being similar in age and body weight, showing no difference in physical activity and feeding behaviour when assessed before drug injections. Starting at 13–14 weeks of age, body weights differed between wild-type and *Cnr1*$^{-/-}$ littermates. Animals from different experimental groups (vehicle versus drug treatment) were randomly allocated within the experimental order. After finishing experiments (data collection), each mouse ID was modified by a second investigator to blind the respective investigator who performed data analyses.

Mice. *Pomc topaz* (*Pomc-GFP*) transgenic mice were provided by J. Friedman and generated as described earlier[48]. The *Npy-hrGFP* line (strain B6.FVB-Tg (Npy-hrGFP) 1Lowl/J) and *Pomc-Cre* line (strain Tg (Pomc1-cre) 16Lowl/J) were purchased from The Jackson Laboratories. *Pomc-GFP* and *Npy-hrGFP* lines are maintained on a C57BL6 background; the *Pomc-Cre* line is maintained on a mixed background in our laboratory. Note that *Npy-hrGFP* mice allow visualization of AgRP neurons through co-expression of NPY and AgRP in these cells. The first CB_1R knockout (*Cnr1*$^{-/-}$) line used here, as maintained in a C57BL6 background, was generated (sponsored by NIMH) and genotyped as previously described[49]. This line was used for feeding studies (Extended Data Fig. 1c) and ELISA experiments (Fig. 3b). For generation of the second *Cnr1*$^{-/-}$ line used here, mice bearing the transcription blocking cassette flanked by *loxP* sites upstream of the *Cnr1* start codon (*Cnr1+/TB-flox*) were crossed with transgenic mice expressing *Pomc-Cre*. Offspring expressing *Cnr1TB/+, Pomc-Cre* were mated with *Cnr1TB/+* to obtain littermate controls (wild-type; *Pomc-Cre::Cnr1+/+* and *Cnr1+/+*), CB_1R knockout controls (*TB-CB$_1$R*; *Cnr1TB/TB*), and mice expressing CB_1R only in POMC neurons (CB1RPOMC; *Pomc-Cre::Cnr1TB/TB*). These mice were used for specification of POMC CB_1R expression and for validation of CB_1R antibody specificity (Extended Data Fig. 5e). The original breeding pairs of the UCP2 knockout (*Ucp2*$^{-/-}$) line were provided by B. Lowell and were generated as reported previously[50]. C57BL6 mice were purchased from The Jackson Laboratories.

Chemicals and antibodies. The selective CB_1R agonist arachidonoyl 2-chloroethylamide (ACEA; Tocris Biosciences) was delivered pre-dissolved in ethanol (5 mg ml^{-1}). For central injection of ACEA, ethanol was evaporated and ACEA was re-dissolved in dimethylsulfoxide (DMSO). The CBR agonist WIN-55,212-2 mesylate (WIN) and the CB_1R inverse agonist rimonabant (RIMO; Cayman Chemical Company) were dissolved in DMSO and aliquots were stored as a stock solution of 10 mg ml^{-1} at −80 °C for 3 months or less. At the day of use, ACEA, WIN or RIMO was mixed with a small drop of Tween 80 (Sigma-Aldrich) before further dilution in saline and applied intraperitoneally to the animals. For the vehicle group, the same volume of ethanol or DMSO was applied compared to ACEA, WIN or RIMO. Clozapine-*N*-oxide (CNO; Enzo Life Sciences) was dissolved in saline (5 mg ml^{-1}) and stored at −80 °C until day of use. The μ-opioid receptor antagonist naloxone (naloxone hydrochloride; Tocris Bioscience) was dissolved in saline (100 mg ml^{-1}) and stored at −80 °C until day of use. Dihydroethidium (DHE; 10× 1 mg; Life Technologies) was stored light protected at −20 °C until use. Hoechst 33342 trihydrochloride, trihydrate (100 mg; Invitrogen) was stored light protected at 4 °C. Virus molecules (4×10^{12} per ml, dialysed in 350 mM NaCl plus 5% D-sorbitol in PBS) containing inhibitory DREADD (rAAV5/EF1α-DIO-hM4D(Gi)-mCherry) or stimulating DREADD (rAAV5/EF1α-DIO-hM3D(Gq)-mCherry) were purchased from Gene Therapy Center Vector Core. Aliquots were stored at −80 °C. The used primary and corresponding secondary antibodies were stored according to manufacturer's instructions. Primary antibodies are summarized in Supplementary Table 1.

Immunofluorescence. After removal of food at 10:00, fed male mice were injected intraperitoneally with vehicle, WIN (1 mg kg^{-1} body weight) or ACEA (1 or 5 mg kg^{-1}) for 90 min. Overnight fasted male mice were injected (intraperitoneally) with vehicle or RIMO (3 mg kg^{-1}) at 10:00 for 90 min. Mice were then anaesthetized by isofluorane and killed by perfusion fixation (4% paraformaldehyde (PFA) in 0.1 M PBS), followed by overnight post-fixation in 4% PFA. After 30 min washing in PBS, brains were cut into 50-μm-thick sections, and 12 sections per mouse containing the ARC were collected. After 15 min in PBS, the sections were incubated in blocking solution (1:20 normal donkey serum in PBS (in case of CB_1R staining using antisera derived from guinea-pig, 1:20 normal goat serum was used), containing 0.2% Triton X-100) for 30 min at room temperature. Primary antibodies (see Supplementary Table 1) were then applied overnight at room temperature. The next day, sections were washed three times (5 min) in PBS and incubated with the respective secondary antibodies for 1 h at room temperature (donkey anti-rabbit IgG fluor 488 or 568 (dilution 1:250, A-11008; or dilution 1:750, A-11036); donkey anti-sheep IgG fluor 488 (dilution 1:250, A-11015); goat anti-guinea-pig IgG fluor 488, 568 or 633 (dilution 1:250; A-11073; dilution 1:750, A-11075; or dilution 1:500, A-21105); goat anti-mouse fluor 633 (dilution 1:500; A-21052); Life Technologies). Finally, the sections were cover-slipped. Double immunofluorescence labelling for cFOS and POMC in *Pomc-Cre* mice started with incubation of sections in blocking solution (see above). The anti-cFOS antibody was incubated at room temperature overnight. The next day, sections were incubated in biotinylated goat anti-rabbit IgG (dilution 1:250; BA-1000, Vector Laboratories) for 2 h at room temperature. After washing in PBS, sections were incubated in avidin, conjugated with Alexa Fluor 488 (dilution 1:2,500, A-21370, Life Technologies) for 1 h at room temperature. After washing in PBS, sections were incubated in unconjugated anti-rabbit IgG (dilution 1:25, 711-005-152, Jackson ImmunoResearch Laboratories) for 2 h at room temperature. After washing in PBS, the sections were incubated with anti-POMC antibody overnight at room temperature. Finally, sections were incubated with goat anti-rabbit Alexa Fluor 633 (dilution 1:750, A-21070, Life Technologies) in PBS for 1 h at room temperature. Confocal laserscanning microscopy was performed using a Zeiss LSM510 Meta, the confocal system equipped with helium neon lasers (543 and 633 nm excitation lines) and an argon laser (488 nm). Quantification of cFOS and pCREB immunoreactivity in *Pomc-GFP* and *Pomc-Cre* mice was performed by counting the percentage of cFOS- and pCREB-positive POMC–GFP or POMC-immunolabelled cells in 10-μm optical, confocal sections obtained from 50-μm-thick vibratome sections. For each mouse, at least 3–4 50-μm-thick sections containing the ARC were analysed.

Immunocytochemistry for electron microscopy. Male C57BL6 mice were anaesthetized and transcardially perfused with 4% PFA and 0.1% glutaraldehyde. After post-fixation overnight, vibratome sections (50 μm) containing the ARC were immunostained. After overnight incubation of primary antibodies (see Supplementary Table 1), sections were washed, incubated with biotin-conjugated donkey anti-rabbit IgG secondary antibody (dilution 1:250, 711-065-152, Jackson ImmunoResearch Laboratories) for 2 h, washed again, put in avidin–biotin complex (ABC; Vector Laboratories), and developed with 3,3-diaminobenzidine (DAB). After washing, sections were incubated in 6 nm colloidal gold-conjugated donkey ant-sheep IgG secondary antibody (dilution 1:50, 713-195-147, Jackson ImmunoResearch Laboratories) for 2 h. Sections were then osmicated (15 min in 1% osmium tetroxide) and dehydrated in increasing ethanol concentrations. During the dehydration, 1% uranyl acetate was added to the 70% ethanol to enhance ultrastructural membrane contrast. Flat embedding in Durcupan followed dehydration. Ultrathin sections were cut on a Leica ultra microtome, collected on Formvar-coated single-slot grids, and analysed with a Tecnai 12 Biotwin electron microscope (FEI).

Semi-quantitative measurements of α-MSH and β-endorphin immunostaining. Analyses of α-MSH and β-endorphin immunosignals in male mice were performed by use of ImageJ 1.48 s software, following the online tutorial for area measurements and particle counting (http://imagej.nih.gov/ij/docs/pdfs/examples.pdf). In brief, confocal stack images (*z*-stacks; thickness/optical section: 6 μm, resolution: 1,024 × 1,024 pixels) of one side of the PVN were obtained from 50-μm-thick vibratome brain sections, using the same magnification and the same confocal microscope setup for all samples. All images from the different experimental groups in one experiment were taken within the same session at the confocal microscope (Zeiss LSM510 Meta). For *z*-projection, five optical sections from the middle area of each brain slice were used. Images were converted into greyscale (8 bit) and binary images were created. For particle analysis, the range of particle size (in pixel units) was set to resemble α-MSH and β-endorphin immunosignals in the PVN and to avoid selection of unspecific background staining. The same range of particle size was used for all samples. Finally, total numbers of selections ('number'), the sum of the size of all selections ('area'; in square pixels) and the sum of the average grey values of all selection ('mean' grey values) were calculated.

Electrophysiology. Coronal hypothalamic slices containing the ARC of male *Pomc-gfp* or *NPY-hrGFP* mice (3–4 weeks old) were prepared 90 min after *in vivo* saline or ACEA (1 mg kg^{-1} body weight) treatment, as described previously[51]. In brief, mice were anaesthetized with isoflurane and decapitated, the brain was rapidly removed and immersed in a cold (4 °C) and oxygenated cutting solution containing (mM): sucrose 220, KCl 2.5, NaH_2PO_4 1.23, $NaHCO_3$ 26, $CaCl_2$ 1, $MgCl_2$ 6 and glucose 10 (pH 7.3 with NaOH). Coronal hypothalamic slices (300 μm thick) were cut with a Leica vibratome after the brain was trimmed to a small tissue block containing the hypothalamus. After preparation, slices were maintained at room temperature (23–25 °C) in a storage chamber in the aCSF (bubbled with 5% CO_2 and 95% O_2) containing (in mM): NaCl 124, KCl 3, $CaCl_2$ 2, $MgCl_2$ 2, NaH_2PO_4 1.23, $NaHCO_3$ 26, glucose 10 (pH 7.4 with NaOH) for recovery and storage. After recovery at room temperature for at least 1 h, slices were transferred to a recording chamber constantly perfused with bath solution (same as the aCSF except

containing 2.5 mM glucose) at a temperature of 33 °C and a perfusion rate of 2 ml min^{-1} for electrophysiological experiments. Whole-cell patch-clamp recording was performed in *Pomc-GFP* and *Npy-hrGFP* neurons under current clamp with the methods previously reported[51]. Spontaneous membrane and action potential was recorded for 10 min in *Pomc-GFP* and *Npy-hrGFP* neurons from both control and ACEA-treated mice.

To test the effects of ACEA on POMC neurons *in vitro*, hypothalamic slices containing the ARC were prepared from naive male *Pomc-GFP* mice. After a stable recording of membrane and action potential under current clamp or membrane current under voltage clamp (at −60 mV), ACEA was applied, in the presence or absence of TTX (0.5 μM), to the recorded cells through bath application for 10 min. The micropipettes (4–6 MΩ) were made of borosilicate glass (World Precision Instruments) with a micropipette puller (Sutter P-97) and back filled with a pipette solution containing (mM): K-gluconate 135, $MgCl_2$ 2, HEPES 10, EGTA 1.1, Mg-ATP 2.5, Na_2-GTP 0.3, and Na_2-phosphocreatin 10 (pH 7.3 with KOH). Both input resistance and series resistance were monitored throughout the experiments and the former was partially compensated. Only recordings with stable series resistance and input resistance were accepted. All data were sampled at 3–10 kHz, filtered at 3 kHz and analysed with an Apple Macintosh computer using AxoGraph X (Kagi).

Behavioural test (open field). A slightly modified experimental regimen according ref. 52 was used. The apparatus consists of a Plexiglas open field (37 × 37 × 37 cm). Male C57BL6 mice habituated for single cage housing and intraperitoneal injection (see 'food intake measurements'), were put in the open field for 90 min after injection with vehicle or ACEA (5 mg kg^{-1} body weight) and locomotor activity was monitored.

Central cannula placement. Anaesthetized (in mg kg^{-1} body weight: intraperitoenal, 30 ketamine, 6 xylazine; subcutaneous; 0.1 buprenorphine) 10-week-old male C57BL6 or 12-week-old male *Pomc-gfp* mice were placed into a stereotaxic apparatus (model 902; David Kopf instruments). The head was cleaned with alcohol pads and incision was made with a sterile scalpel. To place the guide acute cannula (C315GA/SPC, 26 gauge, 8 mm pedestal size, 6.5 mm below pedestal; Plastics One), a single hole was drilled according to the following coordinates: ARC, bregma, anterior–posterior: −1.2 mm, lateral: +0.25 mm; PVN, bregma, anterior–posterior: −0.6 mm, lateral: +0.25 mm). The guided cannula was placed 1 mm dorsal to the region of interest to minimize damage of target area (ARC, bregma, dorsal–ventral: −5.0 mm, PVN, dorsal–ventral: −4.5 mm. The cannula was fixed in place with acrylic dental cement and one anchoring skull screw. After 30 min, a dummy cannula (C315DC/SPCA, cut length, 1 mm extension) was twisted onto the guided cannula. Mice were allowed to recover for 1 week, monitoring food intake and body weight. Mice with abnormal food intake or body weight were excluded. Mice were handled daily and habituated to the injection procedure for 5 days before experimental procedures. At the day of experiment, mice were held gently, while the dummy cannula was removed and the internal cannula (C315IA/SPC, 33 gauge, 8 mm pedestal size, cut length, 1 mm extension) was inserted. The internal cannula was connected to a plastic tube containing a Hamilton syringe at its end, allowing free movement of mice during injection. For PVN, NALO (5 μg per 0.5 μl saline, injection time 1 min) was injected 5 min before peripheral (intraperitoneal) injection of ACEA or WIN. For ARC, ACEA (1 μg per 0.5 μl saline (10% DMSO, 0.2% Tween-80); 1 min) was added. RIMO (1.25 μg per 0.5 μl saline (25% DMSO, 0.2% Tween-80); 1 min) was injected 5 min before peripheral (intraperitoneal) injection of ACEA or WIN. Internal cannula was kept in brain for another minute before replacement and, finally, dummy cannula was twisted onto guided cannulas and food intake was monitored for 1 and 2 h. In previous experiments equal volume injection of the fluorescent dye Hoechst 33342 (1:10 in saline) into the PVN or ARC was identified as an ideal spread of injection. After finishing food intake experiments, Hoechst 33342 was also used to assess correct cannula placement. Mice with missed target areas were excluded.

Stereotaxic virus injection. Animals were injected as described earlier[3,53]. In brief, bilateral virus injections were made into the ARC of anaesthetized (see 'central cannula placement') 12-week-old male *Pomc-Cre* mice, placed into a stereotaxic apparatus (model 902; David Kopf instruments). Viruses (200 nl) containing the inhibitory DREADD (rAAV5/EF1α-DIO-hM4Di-mCherry) or the stimulating DREADD (rAAV5/EF1α-DIO-hM3Dq-mCherry) were applied into each hemisphere (coordinates: bregma, anterior–posterior: −1.2 mm, dorsal–ventral: −5.8 mm, lateral: ±0.3 mm) by using an air pressure system (injection time 5 min). After surgery, mice were allowed to recover for 1 week and were then acclimated for food intake experiments (see 'food intake measurements'). After finishing the feeding experiments, mice were transcardially perfused and 50-μm-thick vibratome sections of the ARC were prepared. Accurate virus injection into the ARC was verified by analysing local mCherry fluorescence. Mice with 'missed' or 'partial' hits were excluded. Specific (virus) expression in ARC POMC neurons was signified by double fluorescence labelling for mCherry and POMC (see 'immunofluorescence').

Food intake measurements. One week before starting experiments, animals were allowed to habituate to single cage housing. Sham injections were performed 2 days before starting the experiment. The day before experiment, single home cages were changed to avoid mice eating food that may have been deposited in the bedding of the cages. To generally minimize disposal of food, the animals were adapted to lower amounts of food (15–25 g per cage) at starting adaption to single housing. To analyse the effect of the CB_1R agonist ACEA or the CBR agonist WIN on food intake in fed animals, food was removed, new pre-weighed food pellets were applied and food intake of each single-housed animal was measured 1, 2 and 24 h after intraperitoneal injection of ACEA or WIN, compared to vehicle, respectively (injection time: 09:00). In the DREADD feeding experiments, CNO (0.3 mg kg^{-1} body weight; intraperitoneal) was injected 15 min before cannabinoid (ACEA or WIN) treatment. To investigate whether cannabinoid (ACEA or WIN)-induced feeding is mediated via β-endorphin release, the μ-opioid receptor antagonist naloxone (7.5 mg kg^{-1} body weight; intraperitoneal) was injected 15 min before cannabinoid (ACEA or WIN) administration and food intake was analysed 1 h after injection. To address the question whether CB_1R inverse agonist RIMO reduced fasting-induced refeeding, the single-housed animals were fasted overnight before intraperitoneal injection of 3 mg kg^{-1} RIMO or vehicle (injection time: 09:00). Fifteen minutes after RIMO injection, a pre-weighed amount of food was placed into the cage and the food intake was then calculated for 1, 2 and 24 h. At all times the cages were inspected for food spillage and those mice in cages with visible food deposits in the bedding were excluded from the studies.

Radioimmunoassay. Male C57BL6 mice were killed by decapitation and brains were immediately removed. A block of the hypothalamus was cut and a 2-mm slice of the mediobasal forebrain was made using a Leica VT1000P vibratome. Each slice was first incubated with aCSF in an incubator (37 °C, 95% O_2 and 5% CO_2) for 1 h. This was followed by another incubation in aCSF, followed by incubation in aCSF (control) or in aCSF containing 100 nM ACEA (both groups containing 56 mM KCl) for 45 min. The supernatant was then removed and frozen for future RIA analysis. The α-MSH and β-endorphin radioimmunoassays were performed following instructions of the Phoenix pharmaceutical RIA kit (α-MSH, RK-043-01; β-endorphin, RK-022-33).

ELISA. *Ex vivo* measurements of β-endorphin content in mediobasal hypothalamus derived from male wild-type and corresponding *Cnr1*$^{-/-}$ littermates was assessed by use of a commercially available β-endorphin ELISA kit and in accordance to manufacturer's instructions (MyBioSource, MBS703919). In brief, 30 mg of tissue was rinsed and mechanically homogenized in PBS. After two freeze–thaw cycles, the homogenates were centrifuged for 5 min at 5,000*g* and 4 °C. The supernatant was removed and assayed immediately. In parallel, sample protein concentrations were measured using the BCA kit (Thermo scientific, 23228 and 1859078), and β-endorphin content was calculated as nanograms β-endorphin per milligram protein.

qPCR and western blot analyses. Food was removed from cages of fed mice. Male C57BL6 mice were then treated with vehicle or ACEA (1 mg kg^{-1} body weight, intraperitoneal) as described above. After 90 min, the mediobasal hypothalamus was rapidly collected and snap frozen in liquid nitrogen. Total RNA was extracted using TRIzol (Life Technologies) according to manufacturer's instructions. cDNA was synthesized by First-Strand cDNA Kit (Amersham Biosciences) following the manufacturer's instructions. Real-time PCR (Light Cycler 480 System, Roche) was performed using TaqMan gene expression assay for *Psck1*, *Pcsk2*, *Ucp2* and *Cnr1* (Mm00479023_m1, 4331182; Mm00500981_m1, 4331182; Mm00627599_m1, 4331182; Mm01212171_s1, 4331182; Life Technologies). β-actin (mouse *Actb*, 4352341; Life Technologies) was used as housekeeping gene.

Micro-punched ARC from fed male C57BL6 mice, either treated with vehicle or ACEA (1 mg kg^{-1} body weight, intraperitoneal) for 90 min after removal of food, were isolated and lysed in modified RIPA buffer (50 mM Tris-HCl, pH 7.5; 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.1% SDS, 1 mM PMSF), supplemented with protease inhibitor cocktail (Roche, 11 836 170 001), on ice for 30 min followed by centrifugation at 14,000 r.p.m. for 15 min. Protein concentrations were measured using the BCA kit (Thermo Scientific). Proteins were resolved on 8% SDS–PAGE and transferred to PVDF membrane (Millipore, IPVH 15150). Membranes were blocked with 5% dry milk in TBS (50 mM Tris-HCl, pH 7.5; 150 mM NaCl) for 1 h and then followed by overnight incubation at 4 °C with first antibodies (see Supplementary Table 1). After three washes with TBST (TBS plus 0.05% Tween 20), membranes were incubated with anti-rabbit IgG conjugated to horseradish peroxidase (Santa Cruz Biotechnology, sc-2004) for 1 h and developed by ECL kit (Thermo Scientific, 32016). Membranes were reused to detect β-actin (see Supplementary Table 1) after stripping with Restore PLUS Western Blot stripping buffer (46430, Thermo Fisher Scientific). The relative protein amount of PC-1, PC-2 and POMC in each sample was calculated from the respective β-actin immunosignal.

ROS measurement. After removal of food at 10:00, fed age- and weight-matched littermates of male and female *Pomc-GFP* mice were injected with vehicle or

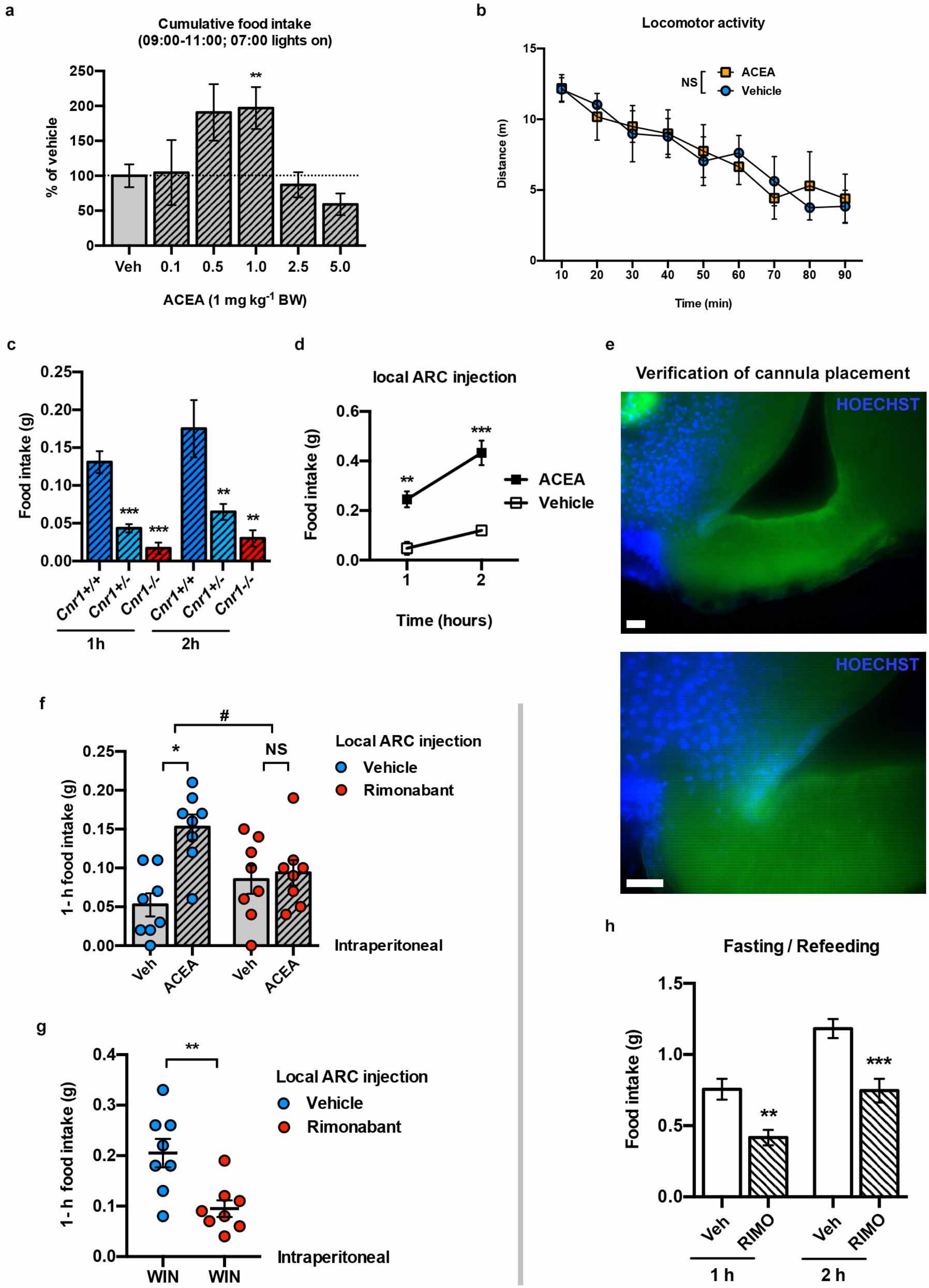
1 mg kg^{-1} ACEA (intraperitoneal) for 90 min. DHE (10 mg) was dissolved in 200 µl DMSO 5 min before injection and kept light-protected at room temperature. Then 240 µl of warm (42 °C) PBS was slowly mixed with 60 µl DHE solution, and 200 µl of the DHE solution (1 mg ml^{-1} in 20% DMSO) was injected via the tail vein 45 min before the animals were anaesthetized by isofluorane and killed by perfusion fixation (4% PFA in 0.1 M PBS). After post-fixation in 4% PFA overnight, the brains were washed in PBS for 30 min and coronal brain sections (50 µm) of the ARC were obtained using a Leica VT1000P vibratome (Leica Microsystems). The slices were mounted on glass slides and coverslipped using DAKO fluorescence mounting medium (Dako North America) and kept at 4 °C for no longer than 1 day. Confocal images were taken by use of an Olympus FV 300 Laserscanning microscope. Semi-quantitative analysis of the DHE signal in *Pomc-GFP* labelled cells was performed by determine the mean dens area of the DHE signal in the respective GFP positive neuron, using imagej 1.48 s software (Rasband, W.S., ImageJ, US National Institutes of Health). To avoid false-positive results, the autofluorescence in animals that were injected with saline instead of DHE was subtracted from the DHE-obtained values.

Brain tissue fractionation. Mouse brain homogenates were purified using a differential centrifugation method described earlier[54]. After removal of food from cages, mice were injected with vehicle or ACEA for 90 min. For isolation of mitochondria-enriched fractions, male and female C57BL6, *Ucp2*$^{-/-}$ and corresponding wild-type littermates were killed and both the hippocampus and the hypothalamus were immediately dissected from brains and put on ice. Tissue was manually homogenized at 4 °C in 10 ml of isolation buffer containing 225 mM mannitol, 75 mM sucrose, 5 mM HEPES, 1 mM EGTA, 1 mg ml^{-1} BSA and 0.3125 mg ml^{-1} protease type VIII (Sigma-Aldrich) at pH 7.4. The homogenate was brought to 20 ml final volume and centrifuged at 2,000*g* for 5 min. The pellet, including the synaptosomal layer was re-suspended in 10 ml of the isolation buffer now containing 0.02% (w/v) digitonin and centrifuged at 12,000*g* for 10 min. The pellet without the synaptosomal layer was re-suspended in 10 ml of isolation buffer and centrifuged at 12,000*g* for 10 min. The pellet was finally re-suspended in 50 µl of a buffer containing 225 mM mannitol, 75 mM sucrose, 5 mM HEPES at pH 7.4.

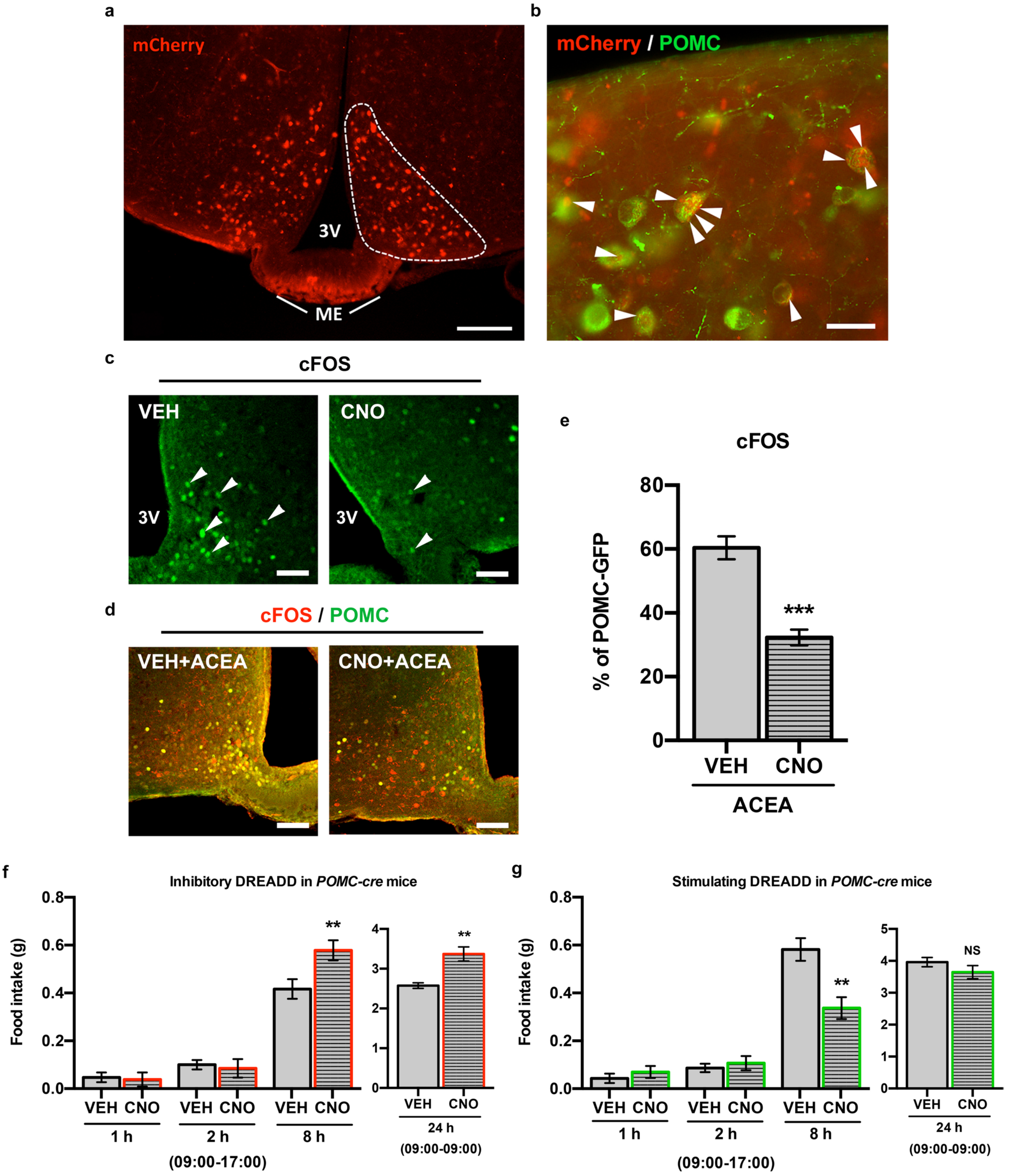
***Ex vivo* mitochondrial respiration.** Protein concentrations of mitochondrial preparations were determined with a BCA protein assay kit (Perbio). For mitochondrial respiration analyses, 0.5 mg ml^{-1} protein was added into the reaction chamber of a Clark type oxygen electrode (Hansatech Instruments) which has been set to 37 °C and filled with 1 ml respiration buffer (100 mM sucrose, 5 mM HEPES, 100 mM KCl, 2 mM KH_2PO_4 and 10 µM EGTA, pH 7.4). Pyruvate (5 mM) and malate (2.5 mM) were then added concomitantly as the oxidative substrates. To determine ADP-dependent (state 3) respiration, ADP (2.5 mM) was added. After the ADP was exhausted, oligomycin (1 µM) was applied to analyse respiration independently of phosphorylation (state 4 activity).

Statistics. Statistical analyses were performed by use of Prism 6.0 software (Graph Pad). $P < 0.05$ was considered statistically significant. *P* values for unpaired comparisons between two groups were calculated by two-tailed Student's *t*-test. One-way ANOVA, followed by Dunnett's multiple comparisons test (all groups against vehicle-treated control group) was used for comparisons between three or more groups. Ordinary two-way ANOVA, followed by Šidák's multiple comparison test was used to examine how food intake was affected by two factors. Interactions between the following treatments were tested: CNO against ACEA or WIN; RIMO against ACEA or WIN; NALO against ACEA or WIN; ACEA treatment against genotype (wild-type littermates versus *Ucp2*$^{-/-}$mice). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Error bars indicate mean ± s.e.m. No statistical methods were used to predetermine sample size.

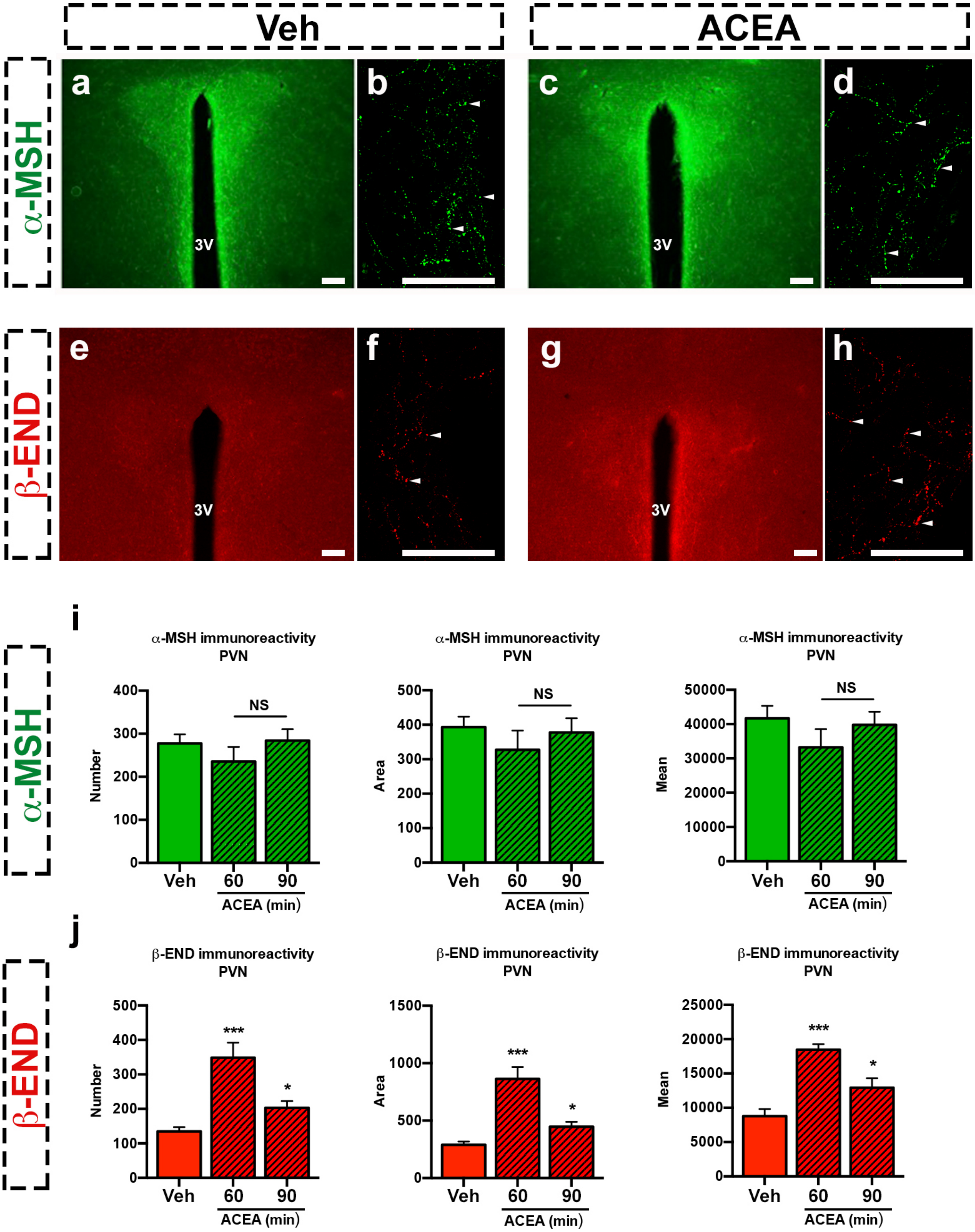
48. Pinto, S. *et al.* Rapid rewiring of arcuate nucleus feeding circuits by leptin. *Science* **304,** 110–115 (2004).
49. Zimmer, A., Zimmer, A. M., Hohmann, A. G., Herkenham, M. & Bonner, T. I. Increased mortality, hypoactivity, and hypoalgesia in cannabinoid CB1 receptor knockout mice. *Proc. Natl Acad. Sci. USA* **96,** 5780–5785 (1999).
50. Zhang, C. Y. *et al.* Uncoupling protein-2 negatively regulates insulin secretion and is a major link between obesity, beta cell dysfunction, and type 2 diabetes. *Cell* **105,** 745–755 (2001).
51. Dietrich, M. O., Liu, Z. W. & Horvath, T. L. Mitochondrial dynamics controlled by mitofusins regulate Agrp neuronal activity and diet-induced obesity. *Cell* **155,** 188–199 (2013).
52. Dietrich, M. O. *et al.* AgRP neurons regulate development of dopamine neuronal plasticity and nonfood-associated behaviors. *Nature Neurosci.* **15,** 1108–1110 (2012).
53. Cetin, A., Komai, S., Eliava, M., Seeburg, P. H. & Osten, P. Stereotaxic gene delivery in the rodent brain. *Nature Protocols* **1,** 3166–3173 (2006).
54. Moreira, P. I. *et al.* Mitochondria from distinct tissues are differently affected by 17β-estradiol and tamoxifen. *J. Steroid Biochem. Mol. Biol.* **123,** 8–16 (2011).


a
Cumulative food intake
(09:00-11:00; 07:00 lights on)
% of vehicle
250
200
150
100
50
0
**
Veh
0.1
0.5
1.0
2.5
5.0
ACEA (1 mg kg-1 BW)
b
Locomotor activity
Distance (m)
15
10
5
0
NS
ACEA
Vehicle
10
20
30
40
50
60
70
80
90
Time (min)
c
Food intake (g)
0.25
0.20
0.15
0.10
0.05
0.00
***
***
**
**
Cnr1+/+
Cnr1+/-
Cnr1-/-
1h
2h
d
local ARC injection
Food intake (g)
0.6
0.4
0.2
0.0
**
***
ACEA
Vehicle
1
2
Time (hours)
e
Verification of cannula placement
HOECHST
HOECHST
f
1- h food intake (g)
0.25
0.20
0.15
0.10
0.05
0.00
#
*
NS
Local ARC injection
Vehicle
Rimonabant
Veh
ACEA
Intraperitoneal
g
1- h food intake (g)
0.4
0.3
0.2
0.1
0.0
**
Local ARC injection
Vehicle
Rimonabant
WIN
WIN
Intraperitoneal
h
Fasting / Refeeding
Food intake (g)
1.5
1.0
0.5
0.0
**
***
Veh
RIMO
1 h
2 h

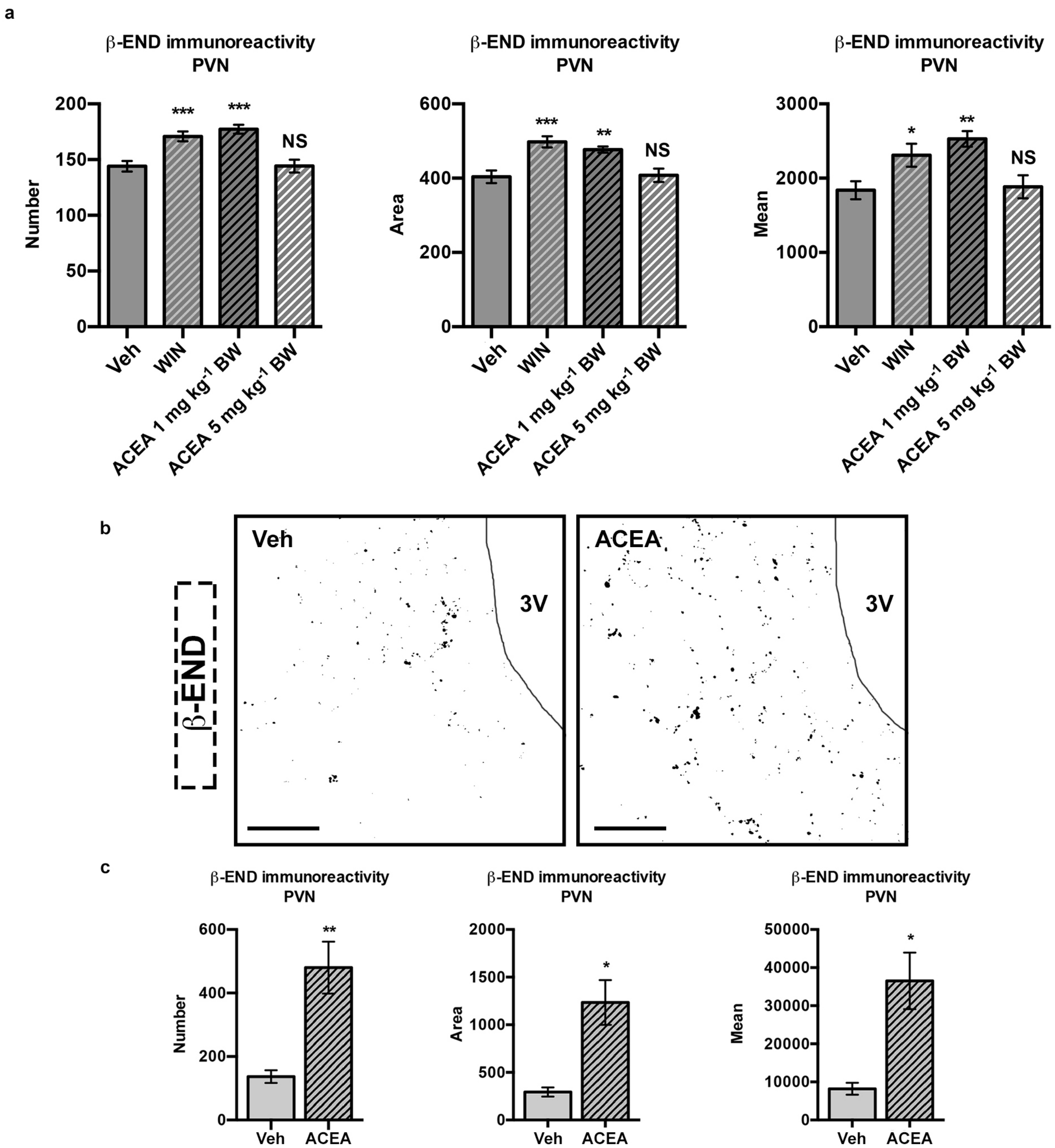
Extended Data Figure 1 | Characterization of CB_1R-dependent food intake. **a**, Bimodal effects of different ACEA doses on food intake in fed mice (vehicle, $n = 23$ mice, $100 \pm 16.3\%$; ACEA (in mg kg^{-1} body weight, intraperitoneal): 0.1, $n = 8$, $104.5 \pm 46.6\%$; 0.5, $n = 3$, $190.8 \pm 40.4\%$; 1.0, $n = 19$, $196.7 \pm 30\%$; 2.5, $n = 16$, $87.1 \pm 18\%$; 5.0, $n = 11$, $59.2 \pm 15.5\%$; $**P < 0.01$ versus vehicle, one-way ANOVA, followed by Dunnett's multiple comparisons test; six independent experiments with litters from different parents). **b**, Neutral dose of ACEA on feeding (5 mg kg^{-1} body weight, intraperitoneal) did not alter locomotor activity of fed mice ($n = 3$ mice/group; $P > 0.05$). **c**, Impaired feeding response to ACEA (1 mg kg^{-1} body weight, intraperitoneal) in CB_1R-heterozygote mice ($Cnr1^{+/-}$, $n = 6$ mice, 1 h: 0.04 ± 0.01 g, 2 h: 0.07 ± 0.01 g) and CB_1R-deficient mice ($Cnr1^{-/-}$, 1 h: $n = 6$, 0.02 ± 0.01 g, 2 h: $n = 4$, 0.03 ± 0.01 g) mice, when compared to CB_1R wild-type mice ($Cnr1^{+/+}$, 1 h: $n = 12$, 0.13 ± 0.01 g, 2 h: $n = 4$, 0.18 ± 0.04 g; $**P < 0.01$, $***P < 0.001$ versus wild-type; two independent experiments). **d**, Central, local ACEA injection into the ARC induced food intake (vehicle, $n = 4$ mice, 1 h: 0.05 ± 0.03 g, 2 h: 0.12 ± 0.01 g; ACEA, $n = 4$, 1 h: 0.25 ± 0.03 g; 2 h: 0.43 ± 0.05 g; $**P < 0.01$, $***P < 0.001$). **e**, Verification of correct ARC cannula placement by HOECHST (blue) injection (representative image (two different magnifications) of four independent experiments). **f**, Hyperphagic CB_1R activation (1 mg kg^{-1} body weight ACEA, intraperitoneal) was abolished by central, local ARC RIMO-mediated CB_1R blockade (vehicle plus vehicle, $n = 8$ mice, 0.05 ± 0.01 g; vehicle plus ACEA, $n = 8$, 0.15 ± 0.02 g; RIMO plus vehicle, $n = 8$, 0.09 ± 0.02 g; RIMO plus ACEA, $n = 8$, 0.09 ± 0.02 g; $*P < 0.05$, $\#P < 0.05$ for interaction between RIMO and ACEA, two-way ANOVA, followed by Šidák's multiple comparisons test; two independent experiments). **g**, Hyperphagic CB_1R activation (1 mg kg^{-1} body weight WIN, intraperitoneal) was reduced by local ARC RIMO-mediated CB_1R blockade (vehicle plus WIN, $n = 8$ mice, 0.21 ± 0.03 g; RIMO+WIN, $n = 8$, 0.1 ± 0.02 g; $**P < 0.01$). **h**, RIMO-induced hypophagic blockade of CB_1R in fasted mice (vehicle, $n = 10$ mice, 1 h: 0.76 ± 0.07 g, 2 h: 1.18 ± 0.07 g; RIMO, $n = 11$ mice, 1 h: 0.42 ± 0.05 g, 2 h: 0.75 ± 0.08 g; $**P < 0.01$, $***P < 0.001$; two independent experiments). Values (biological replicates) denote mean ± s.e.m. If not otherwise stated, P values (unpaired comparisons) by two-tailed Student's t-test. Scale bars, 25 μm.



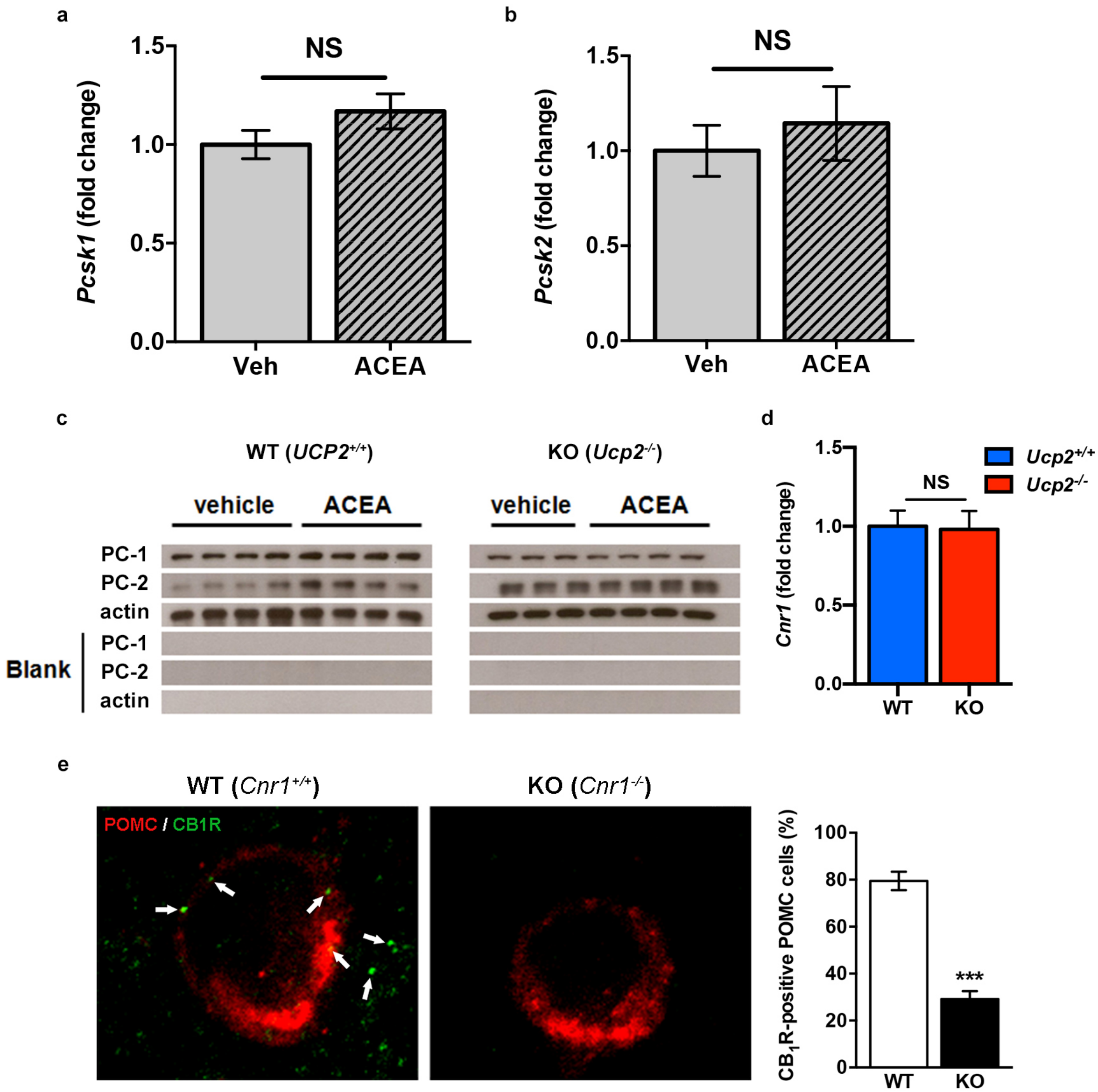
Extended Data Figure 2 | DREADD-mediated regulation of POMC neurons. **a**, Selective DREADD expression specified by local ARC mCherry fluorescence. **b**, POMC neurons (green) contain mCherry-labelled DREADD (red, arrowheads). **c**, CNO-activated inhibitory DREADD reduced ARC cFOS immunolabelled neurons in fed mice (arrowheads). Representative images of four independent experiments (**a**–**c**). **d**, **e**, CNO-activated inhibitory DREADD blocked ACEA-induced POMC activation (cFOS; vehicle plus ACEA, $n = 6$ mice, 60.4 ± 3.6%; CNO plus ACEA, $n = 5$, 32.3 ± 2.5%; $***P < 0.001$). **f**, CNO-activated POMC-specific inhibitory DREADD did not acutely affect feeding but enhanced it after 8 h (vehicle, $n = 17$ mice, 0.42 ± 0.04 g; CNO, $n = 16$, 0.58 ± 0.04 g; 24 h after injection: vehicle, $n = 5$ mice, 2.57 ± 0.07 g; CNO, $n = 5$, 3.37 ± 0.18 g; $**P < 0.01$ versus vehicle; three independent experiments). **g**, CNO-activated POMC-specific stimulating DREADD did not acutely affect feeding but reduced it after 8 h (vehicle, $n = 6$ mice, 0.58 ± 0.05 g; CNO, $n = 6$, 0.34 ± 0.05 g; $**P < 0.01$ versus vehicle; 24 h after injection: vehicle, 3.96 ± 0.15 g; CNO, 3.65 ± 0.21 g; $P > 0.05$ versus vehicle). Values (biological replicates) denote mean ± s.e.m. If not otherwise stated, P values (unpaired comparisons) by two-tailed Student's t-test. Scale bars, 100 μm (**a**), 25 μm (**b**) and 50 μm (**c**, **d**).



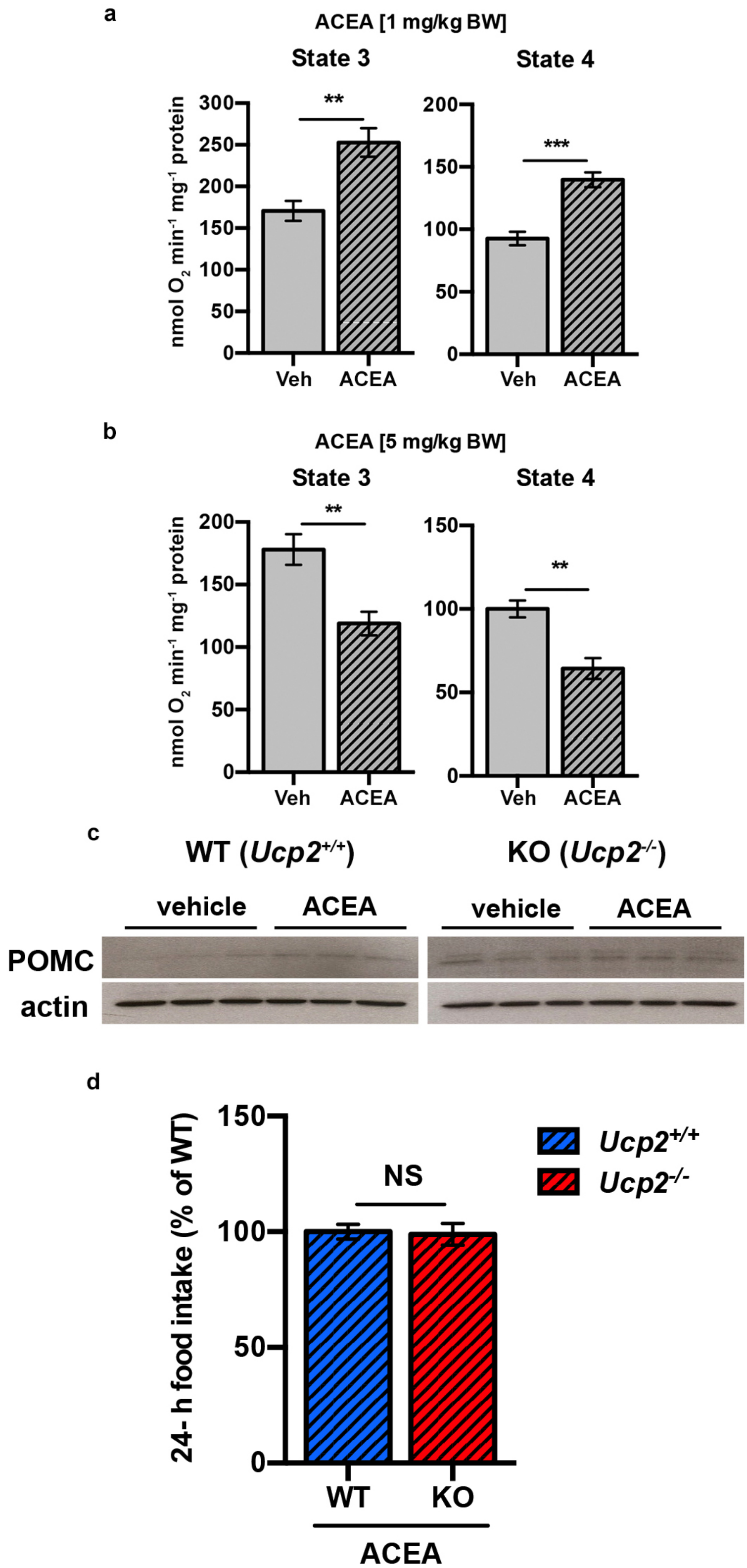
Extended Data Figure 3 | Hyperphagic CB$_1$R activation selectively increased PVN β-endorphin. **a–d**, **i**, PVN α-MSH remained unchanged after hyperphagic CB$_1$R activation (PVN unilateral analysis; vehicle, $n = 6$ values (technical replicates)/6 sections/3 mice (biological replicates); 60 min ACEA, $n = 10/10/5$; 90 min ACEA, $n = 6/6/3$; values, see Extended Data Table 1a). **e–h**, **j**, In contrast, hyperphagic ACEA increased PVN β-endorphin 60 and 90 min after application (PVN unilateral analysis; vehicle, $n = 13$ values/13 sections/6 mice; 60 min ACEA, $n = 4/4/4$; 90 min ACEA, $n = 14/14/7$; values, see Extended Data Table 1b. ***$P < 0.001$, *$P < 0.05$ versus vehicle, one-way ANOVA, followed by Dunnett's multiple comparisons test, two independent experiments using litters from different parents). Error bars indicate mean ± s.e.m. Scale bars, 25 μm.



Extended Data Figure 4 | Bimodal character of ARC CB_1R-driven β-endorphin increase. a, Compared to vehicle (bilateral PVN analysis; $n = 22$ values (technical replicates)/11 sections/4 mice (biological replicates), hyperphagic doses (1 mg kg^{-1} body weight, respectively) of WIN ($n = 24/12/4$) or ACEA ($n = 18/9/3$) induced PVN β-endorphin immunoreactivity. Neutral dose (5 mg kg^{-1} BW) of ACEA ($n = 18/9/3$) on feeding showed no effects (see Extended Data Table 2 for all values). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ versus vehicle, one-way ANOVA, followed by Dunnett's multiple comparisons test. **b**, Representative binary images of four independent experiments showing β-endorphin immunoreactivity after thresholding (image segmentation) using ImageJ software (see Methods). **c**, Compared to vehicle (unilateral PVN analysis; $n = 4$ mice (biological replicates), 2–3 sections (technical replicates) per mouse), central, hyperphagic local ARC injection of ACEA ($n = 5$ mice, 3 sections per mouse) increased PVN β-endorphin immunoreactivity (see Extended Data Table 3 for all values; $^*P < 0.05$, $^{**}P < 0.01$). Error bars indicate mean ± s.e.m. If not otherwise stated, P values (unpaired comparisons) by two-tailed Student's t-test. Scale bars, 100 μm.



Extended Data Figure 5 | Post-transcriptional regulation of hypothalamic pro-protein convertases, normal *Cnr1* expression in *Ucp2*$^{-/-}$ mice and presence of CB$_1$R in POMC neurons. **a**, **b**, ACEA did not affect transcripts of pro-protein convertases 1 (*Pcsk1*) and 2 (*Pcsk2*) (in fold change; *Pcsk1*: vehicle, $n = 11$ mice, 1.00 ± 0.07; ACEA, $n = 10$ mice, 1.17 ± 0.09; *Pcsk2*: vehicle, $n = 11$ mice, 1.00 ± 0.13; ACEA, $n = 11$ mice, 1.14 ± 0.19; $P > 0.05$; two independent experiments). **c**, Representative western blot membranes for PC-1 (~80 kilodaltons (kDa)) and PC-2 (~72 kDa) immunolabelling. **d**, Equal *Cnr1* expression in wild-type and *Ucp2*$^{-/-}$ mice (in fold change: all groups $n = 6$ mice; wild type, 1.00 ± 0.1; *Ucp2*$^{-/-}$, 0.98 ± 0.12; $P > 0.05$). **e**, We have previously shown that antibodies raised against CB$_1$R also recognized the mitochondrial protein, stomatin-like protein 2 (ref. 21). In line with this, mitochondrial labelling of CB$_1$R was found substantially diminished but not completely eliminated in CB$_1$R-KO (*Cnr1*$^{-/-}$) mice[23–25]. We observed that in contrast to wild-type animals (*Cnr1*$^{+/+}$ mice), which showed ~80% (77 out of 97, 79.5 ± 3.9%) of POMC neurons (red fluorescence) to contain labelling with the CB$_1$R antisera (green fluorescence), in CB$_1$R knockout (KO; *Cnr1*$^{-/-}$) mice, less than 30% (37 out of 128, 29.2 ± 3.3%) of POMC neurons retained immunolabelling. Thus, we concluded that a large population of POMC neurons contains CB$_1$R (***$P < 0.001$). All values (biological replicates: **a**–**c**, **d**; biological replicates including technical replicates: **e**) denote mean ± s.e.m. If not otherwise stated, P values (unpaired comparisons) by two-tailed Student's t-test. Scale bar, 25 μm.



Extended Data Figure 6 | Bimodal CB_1R-dependent regulation of mitochondrial respiration and UCP2-dependent control of POMC. **a**, **b**, Bimodal CB_1R-controlled mitochondrial respiration in hippocampus. **a**, Hyperphagic (1 mg kg^{-1} body weight ACEA, intraperitoneal) CB_1R activation increased *ex vivo* mitochondrial respiration (in nmol O_2 min^{-1} mg^{-1} protein; state 3: vehicle, $n = 6$ mice, 170.7 ± 12; ACEA, $n = 8$, 252.7 ± 17.2; state 4: vehicle, 92.7 ± 5.4; ACEA, 139.7 ± 6; $^{**}P < 0.01$, $^{***}P < 0.001$). **b**, Neutral dose of ACEA on feeding (5 mg kg^{-1} body weight, intraperitoneal) reduced mitochondrial respiration (state 3: vehicle, $n = 7$ mice, 178.2 ± 12.2; ACEA, $n = 5$, 118.9 ± 9.4; state 4: vehicle, 100 ± 5.1; ACEA, 64.3 ± 6.3; two independent experiments). **c**, Representative western blot membranes for POMC (pre-POMC, ~31 kDa; POMC, ~27 kDa). **d**, The 24-h food intake did not differ between wild-type ($n = 28$ mice, $100 \pm 3.2\%$) and $Ucp2^{-/-}$ ($n = 29$, $98.9 \pm 4.7\%$; $P > 0.05$) mice after ACEA (1 mg kg^{-1} body weight, intraperitoneal) treatment (six independent experiments using litters from different parents). All values (biological replicates) denote ± s.e.m. If not otherwise stated, P values (unpaired comparisons) by two-tailed Student's t-test.

Extended Data Table 1 | Semi-quantitative measurements of α-MSH and β-endorphin immunoreactivity

a

α-MSH in PVN	Number	Area	Mean
Veh	277.5±21	392.9±30.4	41691±3627
60 min ACEA	235.5±34	327.1±56	33244±5291
90 min ACEA	284±26.5	377.6±41.3	39751±3861

b

β-END in PVN	Number	Area	Mean
Veh	134.9±12.4	289.9±28.2	8774±1048
60 min ACEA	348.5±44***	863.2±103.6***	18468±813.5***
90 min ACEA	203.1±19.5*	446.8±42.4*	12935±1362*

a, α-MSH immunoreactivity; see Extended Data Fig. 3a–d, i. **b**, β-endorphin immunoreactivity; see Extended Data Fig. 3e–h, j. $^{**}P < 0.01$, $^{***}P < 0.001$ versus vehicle.

Extended Data Table 2 | Semi-quantitative measurements of β-endorphin immunoreactivity

β-END in PVN	Number	Area	Mean
Veh	144±4.8	403.7±17.2	1838±121.4
90 min WIN (1 mg kg^{-1} BW)	170.8±4.4***	497.6±15.1***	2308±154.3
90 min ACEA (1 mg kg^{-1} BW)	177.3±3.9***	477±8.4**	2528±104.2
90 min ACEA (5 mg kg^{-1} BW)	144.2±5.7	407.9±17.9	1885±154.8

See Extended Data Fig. 4a. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ versus vehicle.

Extended Data Table 3 | Semi-quantitative measurements of β-endorphin immunoreactivity

β-END in PVN	Number	Area	Mean
Veh	136.7±19.9	294.3. ±47.7	8205±1550
60 min ACEA	480±81.6**	1235±235*	36526±7431*

See Extended Data Fig. 4c. *$P < 0.05$, **$P < 0.01$ versus vehicle.

Extended Data Table 4 | Semi-quantitative measurements of β-endorphin immunoreactivity

β-END in PVN	Number	Area	Mean
WT - Veh	254.5±13.2	287.1±14.8	6885±352.7
WT - 90 min ACEA	311.9±22.6*	365.6±24.1**	8007±407.4*
Ucp2$^{-/-}$ - Veh	247.4±21.3	296±25.2	6840±478.3
Ucp2$^{-/-}$ -90 min ACEA	194.7±14.6*	224±16.9*	4849±328.2**

See Fig. 5j. *$P < 0.05$, **$P < 0.01$ versus wild-type-vehicle or *Ucp2*$^{-/-}$-vehicle, respectively.

doi:10.1038/nature14178

A motor cortex circuit for motor planning and movement

Nuo Li[1], Tsai-Wen Chen[1], Zengcai V. Guo[1], Charles R. Gerfen[2] & Karel Svoboda[1]

Activity in motor cortex predicts specific movements seconds before they occur, but how this preparatory activity relates to upcoming movements is obscure. We dissected the conversion of preparatory activity to movement within a structured motor cortex circuit. An anterior lateral region of the mouse cortex (a possible homologue of premotor cortex in primates) contains equal proportions of intermingled neurons predicting ipsi- or contralateral movements, yet unilateral inactivation of this cortical region during movement planning disrupts contralateral movements. Using cell-type-specific electrophysiology, cellular imaging and optogenetic perturbation, we show that layer 5 neurons projecting within the cortex have unbiased laterality. Activity with a contralateral population bias arises specifically in layer 5 neurons projecting to the brainstem, and only late during movement planning. These results reveal the transformation of distributed preparatory activity into movement commands within hierarchically organized cortical circuits.

The motor cortex is critical for planning and the execution of voluntary movements[1]. Unilateral lesions in premotor areas of motor cortex disrupt planning of movements into the contra-lesional space[2–4]. Neurons in premotor cortex have activity anticipating specific movements long before movement onset[1,5–8], a neural correlate of movement planning[1]. But intermingled motor cortex neurons show puzzlingly diverse selectivity[9,10] for multiple movement directions with complex dynamics[4,11,12]. The relationship of this complex preparatory activity to future movements is not understood. A key question is how preparatory activity evolves into commands that descend to motor centres to trigger movement.

In the mouse, the anterior lateral motor cortex (ALM) is involved in planning directed licking[13]. Unilateral inactivation of ALM during movement planning interferes with upcoming tongue movements in the contralateral direction without impairing movements[13]. A large proportion of ALM neurons exhibit preparatory activity that predicts movements[13], similar to premotor cortex in non-human primates[1,7,8,11,14]. Despite the lateralized deficit from ALM inactivation, ALM neurons in each hemisphere have a preference for contra- or ipsilateral movements in roughly equal proportions[13].

To determine how silencing a brain area with non-lateralized selectivity causes a directional movement bias we measured neuronal activity within hierarchically organized ALM circuits. ALM projection neurons include two major classes: intratelencephalic neurons that project to other cortical areas and pyramidal tract neurons that project out of the cortex, including to motor-related areas in the brainstem[15,16]. Intratelencephalic neurons connect to other intratelencephalic neurons and excite pyramidal tract neurons, but not vice versa. Pyramidal tract neurons are thus at the output end of the local ALM circuit[16–19]. We show that equal proportions of intratelencephalic neurons have preparatory activity for either ipsi- or contralateral movements. Contralateral population activity in pyramidal tract neurons arises late during movement planning to drive directional licking. Our results reveal the flow of information within motor cortex circuits involved in converting preparatory activity into movements.

ALM is required for movement planning

We tested head-fixed mice ($n = 42$) in a whisker-based object location discrimination task[13,20,21]. In each trial, a vertical pole was presented in one of two positions (anterior or posterior) during a sample epoch (1.3 s) (Fig. 1a, b). Mice discriminated pole location using their whiskers. During a subsequent delay epoch (1.3 s) mice planned the upcoming response. An auditory 'go' cue (0.1 s) signalled the beginning of the response epoch, when mice reported the perceived pole position by licking one of two 'lickports' (posterior, lick right; anterior, lick left) (mean per cent correct, 78.4%; responses before the go cue, ~13%).

Unilateral photoinhibition (Methods) of anterior lateral motor cortex (ALM, centred at 2.5 mm anterior, 1.5 mm lateral), limited to the delay epoch, produced an ipsilateral bias (Fig. 1c, d)[13]. Inhibiting left ALM caused a tendency to lick the left port, resulting in increased performance in 'lick left' trials and decreased performance in 'lick right' trials (Fig. 1d). The reversed pattern of bias was observed when inhibiting right ALM. This ipsilateral bias was smaller when photoinhibiting during the sample epoch. ALM activity during the delay period is required for movement planning.

ALM contains neurons with bilateral selectivity

We recorded single units ($n = 1,408$; 19 mice) from left ALM in mice performing object location discrimination. Here we focus on putative pyramidal neurons (Extended Data Fig. 1; $n = 1,245$, Methods) because they carry signals out of the motor cortex to cause movements. A large fraction (73%, 912/1,245) of neurons distinguished trial types ($P < 0.05$, two-tailed t-test, not corrected for multiple comparisons) (Fig. 2b)[13]. Selectivity emerged in the sample epoch, increased throughout the delay epoch, and reached a maximum at the beginning of the response epoch (Fig. 2c and Extended Data Fig. 1d, f). Individual neuron responses were dynamic, despite stable selectivity at the level of the population (Fig. 2c, bottom panel and Extended Data Fig. 2a). ALM neuron responses were diverse: subsets of neurons showed selective preparatory activity (Fig. 2b, left column; Fig. 2d, 219/912 neurons), peri-movement activity during the response epoch (Fig. 2b, right column; 389/912), or both (Fig. 2b, middle column; 304/912). On error trials, when mice licked in the opposite direction to the instruction provided by object location (Fig. 1a), neuronal activity predicted the licking direction (Extended Data Fig. 2b, c)[13]. Such choice-specific activity is consistent with a role in planning and driving movements.

Unilateral inactivation of preparatory activity during the delay epoch caused an ipsilateral bias (contra-lesional deficit) in the upcoming

[1]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA. [2]Laboratory of Systems Neuroscience, National Institute of Mental Health, Bethesda, Maryland 20892, USA.

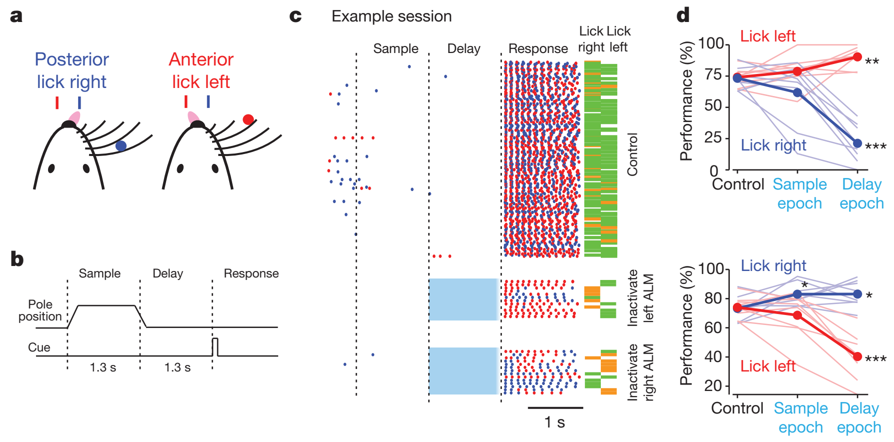


Figure 1 | ALM is required for movement planning. **a**, Head-fixed mouse responding lick right or lick left based on pole location. **b**, The pole was within reach during the sample epoch. Mice responded with licking after a delay and an auditory go cue. **c**, Behavioural data. Individual licks, dots (blue, lick right; red, lick left). Right, bars indicate the performance (green, correct; orange, incorrect). Cyan region, photoinhibition. **d**, Performance with photoinhibition of left (top) or right (bottom) ALM during different trial epochs. Thick lines, mean; thin lines, individual mice ($n = 8$). $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; one-tailed test, bootstrap (Methods).

movement (Fig. 1d). This suggests that ALM neurons have lateralized preference for contralateral movements (that is, neurons from left ALM respond more during lick right trials). We categorized individual neurons as 'contra-preferring' or 'ipsi-preferring' based on spike counts across the trial epochs. Paradoxically, this simple measure did not detect a preference for contralateral licking in ALM spike rates (Fig. 2c, e). Significant contralateral bias also did not appear in the overall spike count across the recorded population of neurons (Extended Data Fig. 1; $P = 0.13$, two-tailed t-test against 0). Individual neurons were either contra-selective (Fig. 2b, top row), or ipsi-selective (Fig. 2b, middle row), or showed mixed selectivity (Fig. 2b, bottom row). The contra-preferring and ipsi-preferring neurons were present in equal proportions (Fig. 2c, e).

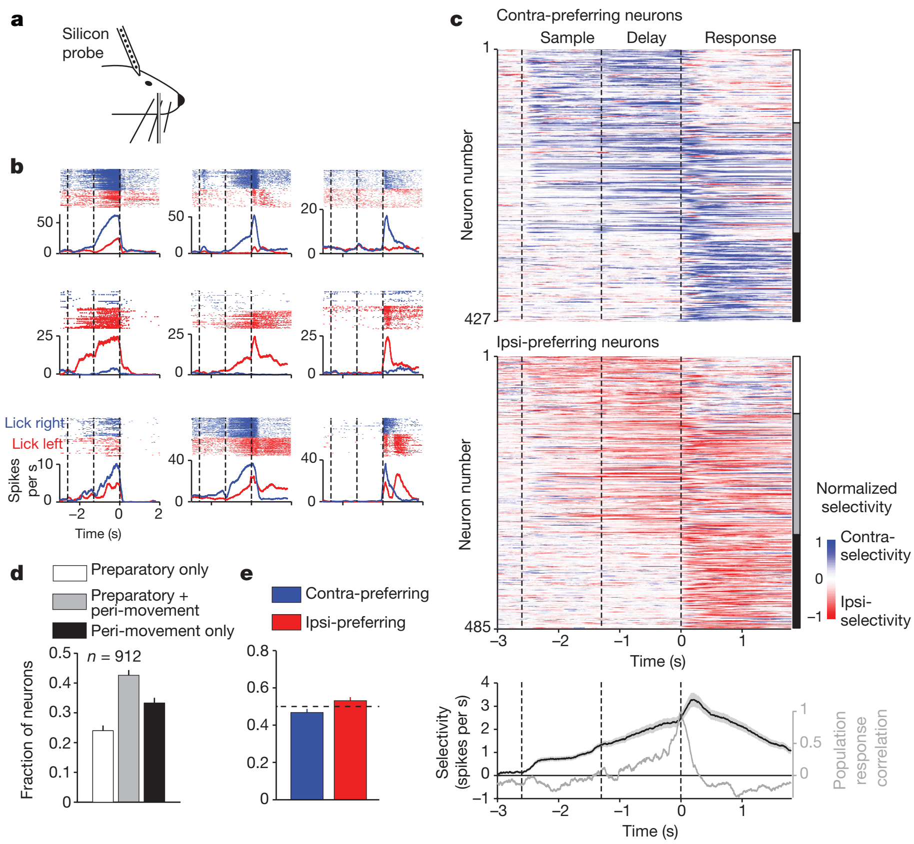


Figure 2 | ALM contains neurons with bilateral movement selectivity. **a**, Silicon probe recordings. **b**, Peri-stimulus time histograms of nine example ALM neurons. Correct lick right (blue) and lick left (red) trials only. Dashed lines, behavioural epochs. **c**, ALM population selectivity. Top, selectivity is the difference in spike rate between the preferred and non-preferred trial type, normalized to the peak selectivity (Methods). Only putative pyramidal neurons with significant trial selectivity are shown ($n = 912/1{,}245$). Bottom, average population selectivity in spike rate (black line, ±s.e.m. across neurons, bootstrap) and population response correlation (grey line; Pearson's correlation between the population response at a particular time and the population response at the onset of the go cue. $t = 0$; Methods). Averaging window, 200 ms. **d**, Proportion of neurons with preparatory and peri-movement activity. **e**, Proportion of contra-preferring versus ipsi-preferring neurons. Error bars, s.e.m. across mice, bootstrap.

Pyramidal tract neurons control directional licking

Selectivity may be distributed non-uniformly across cortical cell types. ALM is interconnected with other cortical areas, including the contralateral ALM (via the corpus callosum), by intratelencephalic neurons (Extended Data Fig. 3a). Intratelencephalic neurons also target ipsi- and contralateral striatum, but avoid subcortical structures involved in the control of movements (Extended Data Fig. 4). Pyramidal tract neurons in layer 5B project to subcortical structures that control movement, including the superior colliculus, brainstem and spinal cord. Pyramidal tract neurons in ALM project to the contralateral intermediate nucleus of the reticular formation, which is presynaptic to the hypoglossal nucleus and the intrinsic and extrinsic muscles of the tongue (Extended Data Fig. 3b, c)[22,23]. Injection of retrograde tracers in the ipsilateral pontine nucleus and the contralateral ALM labelled non-overlapping populations of intratelencephalic neurons and pyramidal tract neurons (Extended Data Fig. 3d) (doubly labelled neurons, $<4\%$, $n = 2$ mice). Intratelencephalic and pyramidal tract neurons were spatially intermingled in layer 5B, but intratelencephalic neurons were also found in superficial layers.

We expressed ChR2 selectively in intratelencephalic or pyramidal tract neurons (Extended Data Fig. 4). Unilateral photostimulation of ALM intratelencephalic or pyramidal tract neurons ('high' power, 0.8 mW, 2 ms pulses at 20 Hz, Methods) triggered contralateral licking with short latencies (Extended Data Fig. 3e), even in untrained mice (Extended Data Fig. 3f)[15,24–26]. This is consistent with the crossed pyramidal tract projection to the brainstem (Extended Data Fig. 3a) and the observation that stimulation of one hypoglossal nerve causes tongue movements in the ipsilateral direction[27]. Photostimulating the vibrissal motor cortex (anterior 1 mm, lateral 1 mm) caused whisker protraction but not licking (Extended Data Fig. 3e–g)[24–26]. Similarly, in mice performing object location discrimination high-power photostimulation of either intratelencephalic or pyramidal tract neurons in ALM produced premature licking in the contralateral direction (Extended Data Fig. 3e). This observation shows that lateralized licking can be triggered by population

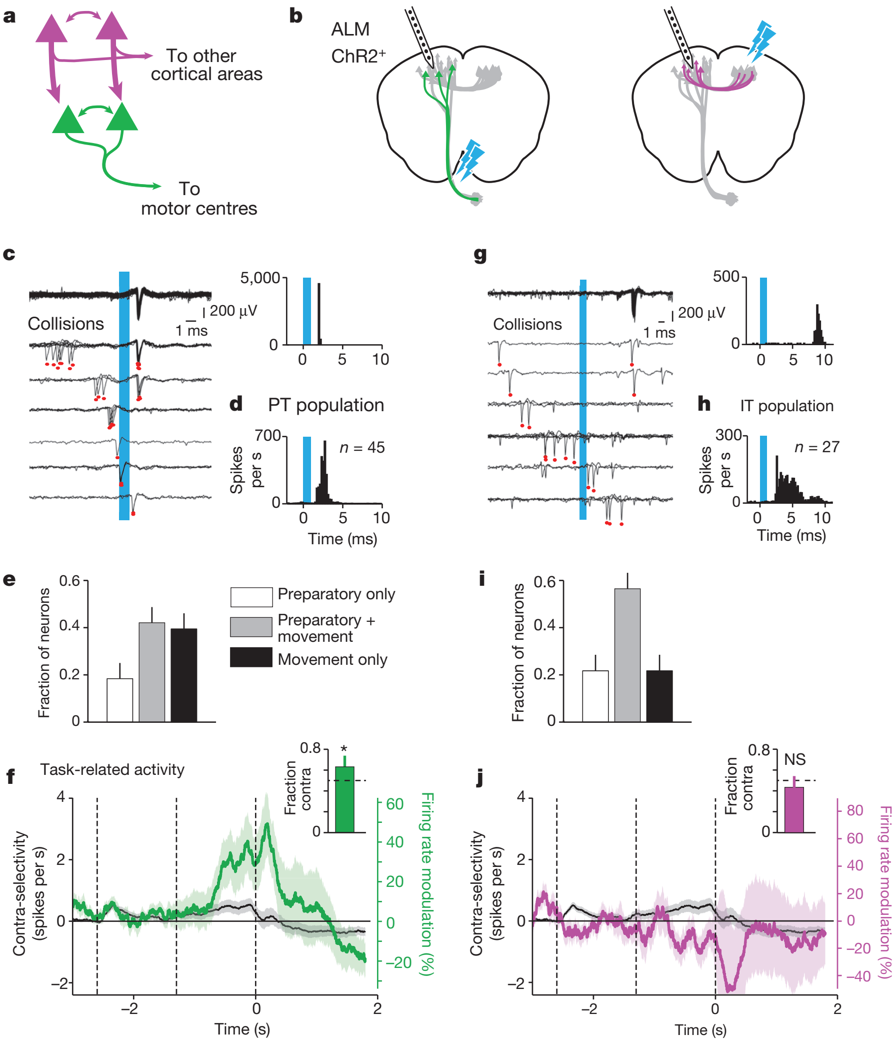


Figure 3 | Cell-type-specific electrophysiology. **a**, ALM circuit involving pyramidal tract (green) and intratelencephalic (magenta) neurons. **b**, Cell-type-specific recordings. Cyan arrows, photostimulation. **c**, Left, recordings from an example pyramidal tract neuron during photostimulation (cyan bar) with collisions. As spontaneous spikes occurred closer in time to the light stimulus, the probability of observing a light-evoked spike decreased (top to bottom traces). Red ticks, individual spikes. Right, peri-stimulus time histogram shows the light-evoked response for the pyramidal tract neuron. **d**, Light-evoked response across all pyramidal tract (PT) neurons ($n = 45$). **e**, Fraction of pyramidal tract neurons sorted by preparatory vs. peri-movement activity. **f**, Contra-selectivity across all pyramidal tract neurons (green, mean ± s.e.m.) and all task-selective neurons (black). Contra-selectivity is the spike rate difference between lick right and light left trials. Insert, fraction of contra-preferring pyramidal tract neurons. $*P < 0.05$, significantly more contra-preferring neurons than ipsi-preferring neurons, one-tailed test, bootstrap. **g–j**, Same as **c–f** for intratelencephalic (IT) neurons. Error bars, s.e.m. across animals, bootstrap.

activity arising in contralateral ALM pyramidal tract neurons. Stimulation of intratelencephalic neurons is also expected to cause lateralized licking because intratelencephalic neurons locally excite pyramidal tract neurons[17].

A movement command emerges in the pyramidal tract neurons

We next recorded selectively from ChR2-expressing intratelencephalic or pyramidal tract neurons during behaviour. During extracellular recordings in left ALM we searched for neurons antidromically activated by photostimulating ('tagging') axonal sites in the pons (pyramidal tract) or contralateral ALM (intratelencephalic)[28] (Methods, Fig. 3b and Extended Data Fig. 5a, c). Photostimulation triggered action potentials in ALM, with short latency and sub-millisecond temporal jitter (mean, 0.15 ms) (Fig. 3c, g and Extended Data Fig. 5b, d). To exclude indirectly activated neurons we tested for collisions between action potentials triggered by axonal photostimulation and somatic action potentials (Methods, Fig. 3c, g and Extended Data Fig. 5b, d)[29,30]. In total, 45 pyramidal tract neurons and 27 intratelencephalic neurons passed the collision test (out of 1,408 neurons recorded in 19 mice) (Methods, Extended Data Fig. 5f). Pyramidal tract neurons had shorter latencies than intratelencephalic neurons (pyramidal tract mean, 2.7 ms; intratelencephalic mean, 4.6 ms) (Extended Data Fig. 5e)[6,31].

Both intratelencephalic and pyramidal tract neurons displayed preparatory and movement-related activity (Fig. 3e, i)[5,6,31]. Contra-preferring neurons outnumbered ipsi-preferring neurons for the pyramidal tract neurons (Fig. 3f; $P < 0.05$, bootstrap, Methods), but not the intratelencephalic neurons (Fig. 3j, $P = 0.74$). Robust contralateral selectivity (the spike count difference for lick right trials and lick left trials) emerged in pyramidal tract neuron responses starting 570 ms before the response epoch (Fig. 3f; 34% spike rate modulation; $P < 0.05$, two-tailed t-test against 0 in spike counts), but not for intratelencephalic neurons (−19%

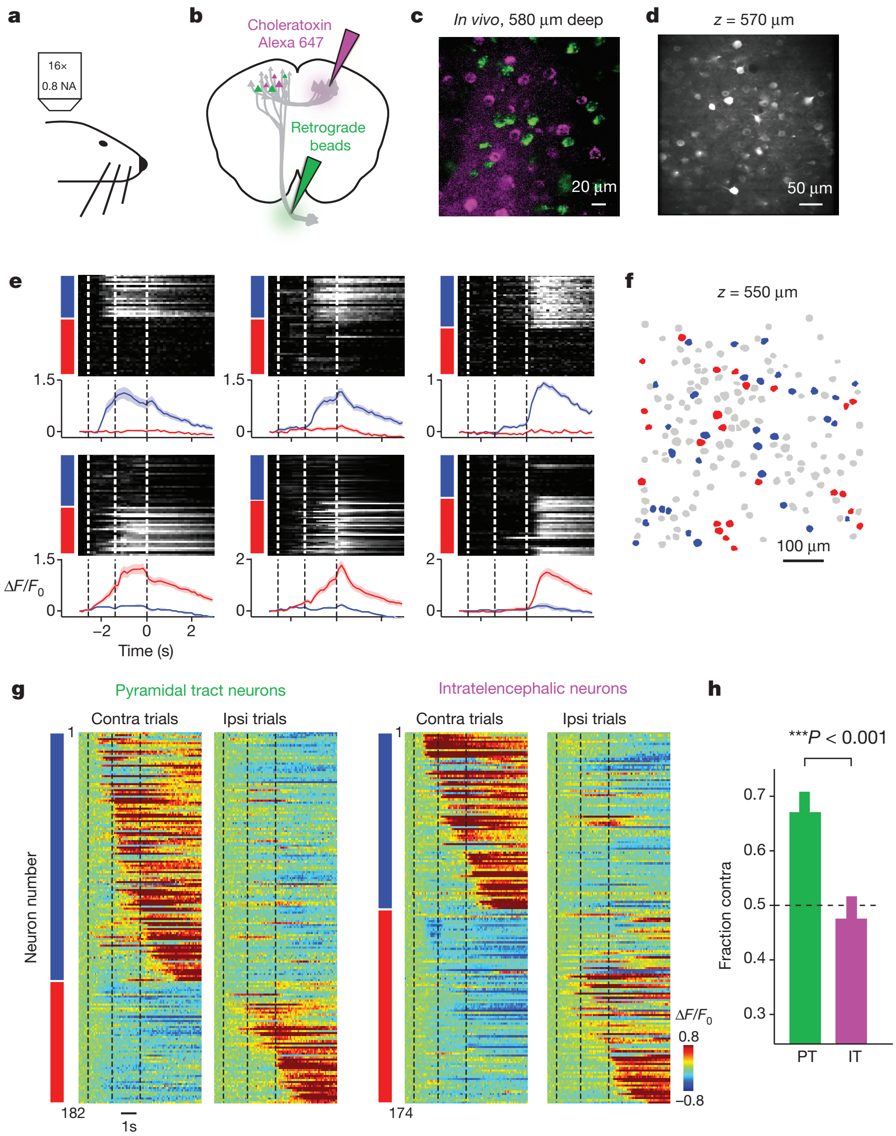


Figure 4 | Cell-type-specific imaging. **a**, Two-photon microscopy. **b**, Labelling pyramidal tract neurons (green) and intratelencephalic neurons (magenta). **c**, Example imaging plane. **d**, Layer 5 neurons expressing GCaMP6s. **e**, Six example ALM neurons. Top, normalized $\Delta F/F_0$ across multiple trials. Coloured bars, trial type (blue, lick right trials; red, lick left trials). Bottom, $\Delta F/F_0$, mean and s.e.m. Dashed lines, behavioural epochs. **f**, Contra-preferring neurons (blue) and ipsi-preferring layer 5 neurons (red) in an example field of view. Grey, non-selective (Methods). **g**, Left, average $\Delta F/F_0$ for selective pyramidal tract neurons. Right, intratelencephalic neurons. Neurons are sorted by their preference (left bar); blue, contra-preferring; red, ipsi-preferring. **h**, Fraction of contra-preferring neurons. Pyramidal tract (PT) neurons were more likely to be contra-preferring than ipsi-preferring ($P < 0.001$, binomial test), but not intratelencephalic (IT) neurons ($P > 0.05$). Pyramidal tract neurons were more likely to be contra-preferring compared to intratelencephalic neurons ($P < 0.001$, χ^2 test). Error bars, s.e.m. across neurons, bootstrap.

spike rate modulation; $P = 0.40$) nor for all task-selective neurons (8% spike rate modulation; $P = 0.13$) (Fig. 3j). Contralateral population selectivity thus arises selectively in pyramidal tract neurons hundreds of milliseconds before movement.

Two-photon calcium imaging provided a much larger sample of pyramidal tract and intratelencephalic neurons. Pyramidal neurons in the left ALM were labelled with GCaMP6s[32]; in addition subsets of pyramidal tract and intratelencephalic neurons were labelled with retrograde fluorescent markers (Methods, Fig. 4b, c)[33,34]. We imaged randomly selected fields of view in ALM (size, 400–600 μm; 59 fields of view; $n = 4$ mice) (Fig. 4d; Methods). Consistent with single-unit recordings, imaging revealed neurons activated at different times during the task (Fig. 4e). We categorized selective neurons into contra-preferring or ipsi-preferring, based on their average calcium signal during the trial epochs (Fig. 4e). Contra-preferring and ipsi-preferring neurons were observed in all fields of view (Fig. 4f; Methods). The probability that nearby neurons share the same selectivity was not different from randomized data ($P = 0.34$, bootstrap, Methods). The distance between nearest neurons with the same selectivity was also not different from chance ($P = 0.61$, bootstrap). Contra-preferring and ipsi-preferring neurons were thus spatially intermingled.

We next analysed pyramidal tract ($n = 534$) and intratelencephalic ($n = 542$) neurons in layer 5 (imaging depth 450–700 μm; $n = 4$ mice) (Fig. 4g). Approximately a quarter of the labelled pyramidal tract (182/534) and intratelencephalic (174/542) neurons were selective (Methods). Consistent with electrophysiological tagging (Fig. 3), we observed more contra-preferring than ipsi-preferring neurons in the pyramidal tract neuron population (67%, 122/182, Fig. 4g, h; $P < 0.001$, binomial test), but not in the intratelencephalic neurons (fraction of contra-preferring neurons: 48%, 83/174; Fig. 4g, h). Contra-lateral preference thus emerges in populations of pyramidal tract neurons.

Pyramidal tract neurons drive upcoming movements

Cell-type-specific recordings revealed that a contralateral selective population signal emerges in pyramidal tract neurons late in the delay epoch, before movement onset (Fig. 3f, selectivity onset: 570 ms before cue; reaction time post cue onset across mice: 61.1 ± 10.5 ms, mean ± s.d.). This suggests that the contra-preferring population activity in pyramidal tract neurons drives the upcoming contralateral licking. We tested this hypothesis directly by activating pyramidal tract neurons during the delay epoch.

We attenuated photostimuli below threshold for activating movement during the sample or delay epochs ('subthreshold' power, <0.25 mW, 2 ms pulses, 20 Hz). Photostimulation was deployed on 25% of the trials either during the sample or delay epochs and ended before the response cue (Methods). Subthreshold photostimulation of the pyramidal tract neurons during the delay epoch biased licking to the contralateral direction. Subthreshold stimulation of pyramidal tract neurons in the left ALM produced a rightward bias, resulting in a decrease in performance in the lick left trials and vice versa (Fig. 5). The effect size was monotonically dependent on laser power (Extended Data Fig. 6b).

Notably, stimulation of pyramidal tract neurons during the sample epoch also caused a significant contralateral bias (Fig. 5b). Simultaneous recordings and optogenetics revealed that subthreshold stimulation of pyramidal tract neurons caused changes in ALM activity and a directional bias that persisted for seconds after cessation of the photostimulus (Extended Data Fig. 7). Activating pyramidal tract neurons is thus sufficient to generate preparatory activity and bias future movements.

Subthreshold photostimulation of the intratelencephalic neurons also caused behavioural changes, but these varied across mice and depended non-monotonically on laser intensity (Extended Data Fig. 6d). This unpredictable and inconsistent effect of intratelencephalic neuron activation on licking direction could be related to the intermingled contralateral and ipsilateral selectivity found in intratelencephalic neurons (Figs 3 and 4). Photostimuli with different powers and in different animals will activate random subsets of intratelencephalic neurons with random net

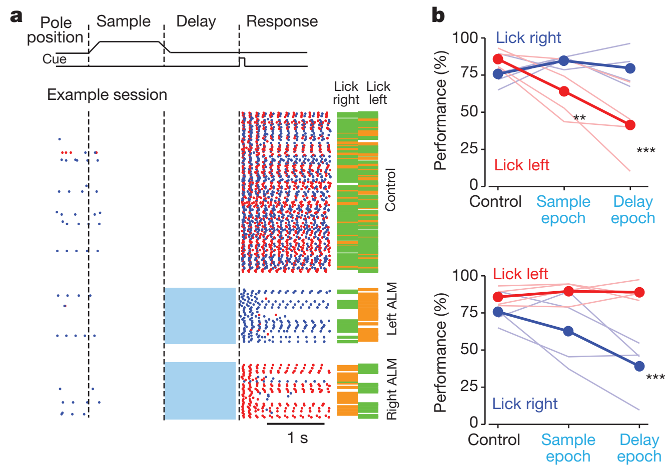


Figure 5 | Preparatory activity in pyramidal tract neurons drives upcoming movements. a, Top, trial structure. Bottom, data from one example session with pyramidal tract neuron activation in the delay epoch. Colour code as Fig. 1c. **b**, Performance in control trials and with pyramidal tract neuron photostimulation in different trial epochs. Top, photostimulation in left ALM. Bottom, right ALM. Colour code as Fig. 1d. $^{**}P < 0.01$; $^{***}P < 0.001$; one-tailed test, bootstrap (Methods). Photostimulation data with multiple subthreshold laser intensities were pooled (0.12–0.8 mW). See Extended Data Fig. 6 for dose response.

selectivity; the resulting network activity will evolve over time to cause unpredictable effects on behaviour (Fig. 2c).

Discussion

Unilateral silencing of ALM during the delay epoch biases future movements in the ipsilateral direction (Fig. 1). ALM also contains a large fraction of neurons with preparatory activity, predicting upcoming movements (Fig. 2)[13], similar to primate premotor cortex[1,5,7,8]. These functional data as well as anatomical studies[35,36] suggest that ALM is homologous to premotor cortex in non-human primates. ALM overlaps with rostral forelimb area, a previously defined premotor region in rats involved in the control of limb movement[36]. Although we studied ALM in the context of licking, ALM and nearby motor cortical areas also have roles in planning other movements[4,37,38] (unpublished data).

Preparatory activity is not an attenuated form of peri-movement activity, but evolves with complex dynamics[1] (Fig. 2b, c and Extended Data Fig. 2a). Lateralized licking is driven by contralateral population activity in populations of pyramidal tract neurons in ALM (Figs 3, 4 and Extended Data Fig. 3). Pyramidal tract neurons are downstream of intratelencephalic neurons, which show contralateral and ipsilateral selectivity with little contralateral bias. This suggests that during movement planning distributed preparatory activity in intratelencephalic neuron networks is converted into a movement command in pyramidal tract neurons ('output-potent' activity)[39], which ultimately triggers directional movements.

Contralateral population activity in pyramidal tract neuron populations still appears hundreds of milliseconds before movement onset (Fig. 3f). Why this pyramidal tract neuron activity, the presumed motor command, does not cause early movements is not understood. One possibility is that a disinhibitory signal from the basal ganglia[40], triggered by the go cue[41], may be necessary to release brainstem motor programs. The same signal could also be routed into ALM through the motor thalamus to reorganize pyramidal tract neuron activity further to make it even more potent for triggering movement[39].

Similar to ALM, human premotor cortex shows bilateral symmetric preparatory potentials (Bereitschaftpotential)[42]. Bilateral coding of movement directions has been observed in single-unit recordings from primate[11,12,43] and rat motor cortex[4]. In contrast to the bilateral selectivity in neurons, unilateral damage to the premotor cortices frequently

causes hemi-neglect, where subjects show deficit in planning movements to the contra-lesional space[2–4]. Our results suggest that the contra-lesional bias in premotor hemi-neglect is caused by damage to the pyramidal tract neurons.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 August; accepted 23 December 2014.
Published online 25 February 2015.

1. Shenoy, K. V., Sahani, M. & Churchland, M. M. Cortical control of arm movements: a dynamical systems perspective. *Annu. Rev. Neurosci.* **36,** 337–359 (2013).
2. Kerkhoff, G. Spatial hemineglect in humans. *Prog. Neurobiol.* **63,** 1–27 (2001).
3. Rizzolatti, G., Matelli, M. & Pavesi, G. Deficits in attention and movement following the removal of postarcuate (area 6) and prearcuate (area 8) cortex in macaque monkeys. *Brain* **106,** 655–673 (1983).
4. Erlich, J. C., Bialek, M. & Brody, C. D. A cortical substrate for memory-guided orienting in the rat. *Neuron* **72,** 330–343 (2011).
5. Tanji, J. & Evarts, E. V. Anticipatory activity of motor cortex neurons in relation to direction of an intended movement. *J. Neurophysiol.* **39,** 1062–1068 (1976).
6. Turner, R. S. & DeLong, M. R. Corticostriatal activity in primary motor cortex of the macaque. *J. Neurosci.* **20,** 7096–7108 (2000).
7. Crutcher, M. D. & Alexander, G. E. Movement-related neuronal activity selectively coding either direction or muscle pattern in three motor areas of the monkey. *J. Neurophysiol.* **64,** 151–163 (1990).
8. Riehle, A. & Requin, J. Monkey primary motor and premotor cortex: single-cell activity related to prior information about direction and extent of an intended movement. *J. Neurophysiol.* **61,** 534–549 (1989).
9. Scott, S. H. Inconvenient truths about neural processing in primary motor cortex. *J. Physiol. (Lond.)* **586,** 1217–1224 (2008).
10. Huber, D. *et al.* Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* **484,** 473–478 (2012).
11. Rizzolatti, G. *et al.* Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Exp. Brain Res.* **71,** 491–507 (1988).
12. Arce, F. I., Lee, J. C., Ross, C. F., Sessle, B. J. & Hatsopoulos, N. G. Directional information from neuronal ensembles in the primate orofacial sensorimotor cortex. *J. Neurophysiol.* **110,** 1357–1369 (2013).
13. Guo, Z. V. *et al.* Flow of cortical activity underlying a tactile decision in mice. *Neuron* **81,** 179–194 (2014).
14. Bruce, C. J. & Goldberg, M. E. Primate frontal eye fields. I. Single neurons discharging before saccades. *J. Neurophysiol.* **53,** 603–635 (1985).
15. Komiyama, T. *et al.* Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature* **464,** 1182–1186 (2010).
16. Shepherd, G. M. Corticostriatal connectivity and its role in disease. *Nature Rev. Neurosci.* **14,** 278–291 (2013).
17. Kiritani, T., Wickersham, I. R., Seung, H. S. & Shepherd, G. M. Hierarchical connectivity and connection-specific dynamics in the corticospinal-corticostriatal microcircuit in mouse motor cortex. *J. Neurosci.* **32,** 4992–5001 (2012).
18. Brown, S. P. & Hestrin, S. Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. *Nature* **457,** 1133–1136 (2009).
19. Morishima, M. & Kawaguchi, Y. Recurrent connection patterns of corticostriatal pyramidal cells in frontal cortex. *J. Neurosci.* **26,** 4394–4405 (2006).
20. O'Connor, D. H. *et al.* Vibrissa-based object localization in head-fixed mice. *J. Neurosci.* **30,** 1947–1967 (2010).
21. Guo, Z. V. *et al.* Procedures for behavioral experiments in head-fixed mice. *PLoS ONE* **9,** e88678 (2014).
22. Travers, J. B., Dinardo, L. A. & Karimnamazi, H. Motor and premotor mechanisms of licking. *Neurosci. Biobehav. Rev.* **21,** 631–647 (1997).
23. Stanek, E. t., Cheng, S., Takatoh, J., Han, B. X. & Wang, F. Monosynaptic premotor circuit tracing reveals neural substrates for oro-motor coordination. *eLife* **3,** e02511 (2014).
24. Li, C. X. & Waters, R. S. Organization of the mouse motor cortex studied by retrograde tracing and intracortical microstimulation (ICMS) mapping. *Can. J. Neurol. Sci.* **18,** 28–38 (1991).
25. Hall, R. D. & Lindholm, E. P. Organization of motor and somatosensory neocortex in the albino rat. *Brain Res.* **66,** 23–38 (1974).
26. Neafsey, E. J. *et al.* The organization of the rat motor cortex: a microstimulation mapping study. *Brain Res.* **396,** 77–96 (1986).
27. Bennett, G. A. & Ramsay, A. J. Experimental studies on the movements of the mammalian tongue. I. Movements of the split tongue (dog). *Anat. Rec.* **79,** 39–51 (1941).
28. Petreanu, L., Huber, D., Sobczyk, A. & Svoboda, K. Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nature Neurosci.* **10,** 663–668 (2007).
29. Beloozerova, I. N., Sirota, M. G. & Swadlow, H. A. Activity of different classes of neurons of the motor cortex during locomotion. *J. Neurosci.* **23,** 1087–1097 (2003).
30. Swadlow, H. A. Efferent neurons and suspected interneurons in motor cortex of the awake rabbit: axonal properties, sensory receptive fields, and subthreshold synaptic inputs. *J. Neurophysiol.* **71,** 437–453 (1994).
31. Bauswein, E., Fromm, C. & Preuss, A. Corticostriatal cells in comparison with pyramidal tract neurons: contrasting properties in the behaving monkey. *Brain Res.* **493,** 198–203 (1989).
32. Chen, T. W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499,** 295–300 (2013).
33. Sato, T. R. & Svoboda, K. The functional properties of barrel cortex neurons projecting to the primary motor cortex. *J. Neurosci.* **30,** 4256–4260 (2010).
34. Yamashita, T. *et al.* Membrane potential dynamics of neocortical projection neurons driving target-specific signals. *Neuron* **80,** 1477–1490 (2013).
35. Reep, R. L., Corwin, J. V., Hashimoto, A. & Watson, R. T. Efferent connections of the rostral portion of medial agranular cortex in rats. *Brain Res. Bull.* **19,** 203–221 (1987).
36. Rouiller, E. M., Moret, V. & Liang, F. Comparison of the connectional properties of the two forelimb areas of the rat sensorimotor cortex: support for the presence of a premotor or supplementary motor cortical area. *Somatosens. Mot. Res.* **10,** 269–289 (1993).
37. Murakami, M., Vicente, M. I., Costa, G. M. & Mainen, Z. F. Neural antecedents of self-initiated actions in secondary motor cortex. *Nature Neurosci.* **17,** 1574–1582 (2014).
38. Sul, J. H., Jo, S., Lee, D. & Jung, M. W. Role of rodent secondary motor cortex in value-based action selection. *Nature Neurosci.* **14,** 1202–1208 (2011).
39. Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Cortical activity in the null space: permitting preparation without movement. *Nature Neurosci.* **17,** 440–448 (2014).
40. Hikosaka, O. & Wurtz, R. H. Modification of saccadic eye movements by GABA-related substances. II. Effects of muscimol in monkey substantia nigra pars reticulata. *J. Neurophysiol.* **53,** 292–308 (1985).
41. Pan, W. X., Brown, J. & Dudman, J. T. Neural signals of extinction in the inhibitory microcircuit of the ventral midbrain. *Nature Neurosci.* **16,** 71–78 (2013).
42. Shibasaki, H. & Hallett, M. What is the Bereitschaftspotential? *Clin. Neurophysiol.* **117,** 2341–2356 (2006).
43. Murray, G. M. & Sessle, B. J. Functional properties of single neurons in the face primary motor cortex of the primate. III. Relations with different directions of trained tongue protrusion. *J. Neurophysiol.* **67,** 775–785 (1992).

Acknowledgements We thank B. Dickson, S. Druckmann, J. Dudman, D. Gutnisky, H. Inagaki, V. Jayaraman, D. O'Connor, S. Peron, T. Sato and G. Shepherd for comments on the manuscript and discussion, L. Walendy and E. Ophir for animal training, S. Michael and A. Hu for histology, T. Harris and B. Barbarits for the silicon probe recording system. This work was funded by the Howard Hughes Medical Institute. N.L. is a Helen Hay Whitney Foundation postdoctoral fellow.

Author Contributions N.L., Z.G. and K.S. initiated this study. N.L. performed electrophysiology and optogenetic experiments. T.W.C. performed imaging. T.W.C., N.L., and C.G. performed anatomical experiments. N.L., T.W.C., K.S. analysed data. N.L. and K.S. wrote the paper, with input from all authors.

Author Information Data have been deposited at https://crcns.org/ and can be accessed at http://dx.doi.org/10.6080/K0RF5RZT. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.S. (svobodak@janelia.hhmi.org).

METHODS

No statistical methods were used to predetermine sample size.

Mice. This study is based on data from 52 mice (age > P60). Five VGAT-ChR2-EYFP mice (Jackson laboratory, JAX Stock#014548)[44] and three PV-ires-Cre[45] crossed to Rosa26-LSL-ReaChR, red-shifted channelrhodopsin reporter mice (JAX 28846)[46], were used for photoinhibition experiments (Fig. 1)[13]. Nine Sim1_KJ18-Cre mice (MMRRC 031742), 3 Rbp4-Cre mice (MMRRC 031125), and 4 Tlx_PL56-Cre mice (MMRRC 036547) were used for electrophysiology experiments (Figs 2 and 3)[47]. Four C57Bl/6Crl mice were used for imaging experiments (Fig. 4). Seven Sim1_KJ18-Cre crossed to Ai32 (Rosa26-ChR2 reporter mice, JAX Stock#012569)[48] mice were used for photoactivation behavioural experiments (Fig. 5 and Extended Data Fig. 3), and one of these mice was also used for electrophysiology. Seven Tlx_PL56-Cre crossed to Ai32 mice were used for photoactivation behaviour experiments and two of these mice were also used for electrophysiology. Two Sim1_KJ18-Cre, 2 Tlx_PL56-Cre, 4 C57B1/6Crl mice, 1 Sim1_KJ18-Cre × Ai32 mouse, and 1 Tlx_PL56-Cre × Ai32 mouse were used for anatomy experiments (Extended Data Figs 3 and 4).

All procedures were in accordance with protocols approved by the Janelia Research Campus Institutional Animal Care and Use Committee. Mice were housed in a 12:12 reverse light:dark cycle and tested during the dark phase. On days not tested, mice received 1 ml of water. On other days, mice were tested in experimental sessions lasting 1 to 2 h where they received all their water (range, 0.5 to 2 ml). If mice did not maintain a stable body weight, they received supplementary water[21]. All surgical procedures were carried out aseptically under 1–2% isofluorane anaesthesia. Buprenorphine HCl (0.1 mg kg^{-1}, intraperitoneal injection; Bedford Laboratories) was used for postoperative analgesia. Ketoprofen (5 mg kg^{-1}, subcutaneous injection; Fort Dodge Animal Health) was used at the time of surgery and postoperatively to reduce inflammation. After the surgery, mice were allowed to recover for at least 3 days with free access to water before water restriction.

Behaviour. Mice were prepared for photostimulation and electrophysiology by implantation of a clear skullcap and a headpost[13]. The scalp and periosteum over the dorsal surface of the skull were removed. A layer of cyanoacrylate adhesive (Krazy glue, Elmer's Products Inc.) was directly applied to the intact skull. A custom made headpost[21] was placed on the skull with its anterior edge aligned with the suture lambda (approximately over cerebellum) and cemented in place using clear dental acrylic (Lang Dental Jet Repair Acrylic; part no. 1223-clear). A thin layer of clear dental acrylic was applied over the cyanoacrylate adhesive covering the entire exposed skull, followed by a thin layer of clear nail polish (Electron Microscopy Sciences, part no. 72180).

The behavioural task has been described in detail[13,21]. The stimulus was a metal pin (0.9 mm in diameter), presented at one of two possible positions (Fig. 1). The two pole positions were 4.29 mm apart along the anterior–posterior axis (40° of whisking angle) and were constant across sessions. The posterior pole position was 5 mm from the whisker pad. A two-spout lickport (4.5 mm between spouts) was used to deliver water rewards and record licks. Mouth movements were monitored using a photodiode and an infrared laser diode to obtain reaction time measurements (Thorlabs). High-speed video (Mikrotron Eosens Camera, Norpix, MC1362) was taken over a 11.4 mm × 15.2 mm region at 1 kHz to track the whiskers.

At the beginning of each trial, the vertical pole moved into the plane within reach of the whiskers (0.2 s travel time). The pole remained within reach for 1 s, after which it was retracted (retraction time 0.2 s). Mice made contacts with the object at both pole positions, typically with a different set of whiskers. The sample epoch is defined as the time between the pole movement onset to 0.1 s after the pole retraction onset (sample epoch, 1.3 s total, Fig. 1b). The delay epoch lasted for another 1.2 s after completion of pole retraction (delay epoch, 1.3 s total, Fig. 1b). An auditory go cue indicated the end of the delay epoch (pure tone, 3.4 kHz, 0.1 s duration). Licking early during the trial was punished by a loud alarm sound (siren buzzer, 0.05 s duration), followed by a brief timeout (1–1.2 s). Licking the correct lickport after the go cue led to a small drop of liquid reward (3 μl). Licking the incorrect lickport triggered a timeout (2–5 s). Trials in which mice did not lick within a 1.5 s window after the go cue were rare and typically occurred at the end of a session.

Viral injection and histology. To characterize BAC Cre mice we injected eGFP (AAV2/1.CAG.EGFP, http://www.addgene.com, plasmid 28014) into one hemisphere and Cre-dependent tdTomato (AAV2/1.CAG.FLEX.tdTomato.WPRE, UPenn Viral Core, AV-1-ALL864) into ALM on the other hemisphere (Extended Data Fig. 3a). The ALM injection coordinate was 2.5 mm anterior to bregma, 1.5 mm lateral[13]. The injection was made through the thinned skull using a custom, piston-based, volumetric injection system. Glass pipettes (Drummond) were pulled and bevelled to a sharp tip (outer diameter of 30 μm). Pipettes were back-filled with mineral oil and front-loaded with viral suspension immediately before injection. Fifty nanolitres were injected 500 and 800 μm deep. Two weeks post injection, mice were perfused transcardially with PBS followed by 4% paraformaldehyde (PFA)/0.1 M PBS. The brains were fixed overnight and transferred to 20% sucrose before sectioning on a freezing microtome. Coronal 50 μm free-floating sections were processed using standard fluorescent immunohistochemical techniques. Slide-mounted sections were imaged with a Zeiss microscope, a 10× objective and a Hamamatsu Orca Flash 4 camera. Each coronal section was made up of 80–200 tiles merged with Neurolucida software[47]. For brainstem injections (Extended Data Fig. 3b, c), 150 nl of a 1:2 mix of AAV2/1.CAG.EGFP and red RetroBeads (Lumafluor) was injected into the intermediate nucleus of the reticular formation (6.65 mm posterior to bregma, 1.25 mm lateral, 4.2–4.5 mm deep) and hypoglossal nucleus, 12N (7.25 mm posterior to bregma, 0 mm lateral, 4–4.25 mm deep). Mice were perfused 4 weeks post injection. The brains sections were imaged on an Olympus Macroscope.

Photostimulation. Light from a 473 nm laser (Laser Quantum, part no. Gem 473) or a 594 nm laser (Cobolt Inc., part no. Colbolt Mambo 100) was controlled by an acousto-optical modulator (AOM; Quanta Tech) and a shutter (Vincent Associates). Photostimulation of ALM was performed through the clear skullcap implant by directing the laser over the skull (beam diameter, 400 μm at 4σ). The light transmission through the intact skull is 50%[13]. Photostimulation was deployed on 25% of the behavioural trials. To prevent the mice from distinguishing photostimulation trials from control trials using visual cues, a 'masking flash' (40 1 ms pulses at 10 Hz) was delivered using 470 nm LEDs (Luxeon Star) near the eyes of the mice. The masking flash began as the pole started to move and continued through the end of the epoch in which photostimulation could occur.

For silencing we stimulated cortical GABAergic neurons in VGAT-ChR2-EYFP mice, or parvalbumin-positive interneurons in PV-ires-Cre mice crossed to reporter mice expressing ReaChR (Fig. 1). The two methods resulted in similar photoinhibition (data not shown). We used 40 Hz photostimulation with a sinusoidal temporal profile (1.5 mW average power) and a 100 ms linear ramp during the laser offset (this reduced rebound neuronal activity)[13]. The photoinhibition silenced a cortical area of 1 mm radius (at half maximum) through all cortical layers. The temporal onset of the photoinhibition reached its peak within 20 ms of light onset. The photoinhibition recovered ~100 ms after the onset of the linear ramp[13]. To silence ALM activity during the sample or delay epochs (Fig. 1) we photostimulated for 1.3 s, including the 100 ms ramp, starting at the beginning of the epoch. Thus, photoinhibition always ended before the response cue.

To activate layer 5 intratelencephalic or pyramidal tract neurons (Fig. 5, Extended Data Figs 6 and 7), we photostimulated the motor cortex in the intratelencephalic or pyramidal tract BAC-Cre driver lines crossed to a Rosa26-ChR2 reporter line (Ai32). During behaviour, we used pulses of light (2 ms pulse duration) at 20 Hz (26 pulses, 1252 ms) and a range of peak powers (1.5, 3, 6, 20 mW). The values reported in the figures indicate the average power (0.06, 0.12, 0.24, 0.8 mW; Fig. 5, Extended Data Figs 3 and 6). In untrained mice, we used 5-ms light pulses and a larger range of peak powers (1.5, 6, 13, 22, 40, 47 mW; average power, 0.15, 0.6, 1.3, 2.2, 4, 4.7 mW; Extended Data Fig. 3).

Electrophysiology. A small craniotomy (diameter, 1 mm) was made over ALM one day before the recording sessions[13]. Extracellular spikes were recorded using NeuroNexus silicon probes (part no. A4x8-5mm-100-200-177). The 32 channel voltage signals were multiplexed, recorded on a PCI6133 board at 312.5 kHz (National instrument), and digitized at 14 bit. The signals were demultiplexed into the 32 voltage traces, sampled at 19531.25 Hz and stored for offline analyses. Three to eight recordings were made from each craniotomy. Recording depth was inferred from manipulator readings[13]. To minimize brain movement, a drop of silicone gel (3-4680, Dow Corning, Midland, MI) was applied over the craniotomy after the electrode was in the tissue. The tissue was allowed to settle for several minutes before the recording started.

To optogenetically tag ALM intratelencephalic and pyramidal tract neurons during recording, we first infected ALM neurons with ChR2. Cre-dependent ChR2 virus (AAV2/5.hSyn1.FLEX.hChR2.tdTomato, http://www.addgene.com, plasmid 41015)[49] was injected into three Rbp4-Cre mice (targeting both layer5 intratelencephalic and pyramidal tract neurons), four Tlx_PL56-Cre mice (layer 5 intratelencephalic neuron), and nine Sim1_KJ18-Cre mice (pyramidal tract neurons). One-hundred-nanolitre volumes were injected 500 and 800 μm deep. We also used two Tlx_PL56-Cre cross Ai32 transgenic mice for antidromic tagging of intratelencephalic neurons. However, expression of ChR2 in large numbers of neurons in the transgenic mice resulted in highly synchronized activity upon light stimulation, and consequently, difficulty in isolating antidromically activated single units. To photostimulate the axons of pyramidal tract neurons, an optical fibre (Thorlabs, part no. CFML12L05) was implanted into the reticular formation based on stereotaxic coordinates (5.5 mm posterior, 1 mm lateral, 5 mm deep), ipsilateral to the viral injection site. One to four months after the infection and fibre implant, silicon probe recordings were made from the virus injection site. For photostimulation of intratelencephalic neuron axons, photostimulation was through a cranial window over the contralateral ALM. Pairs of laser pulses (1 ms duration, 47–82 mW peak power, separated by 10 ms) were deployed every 500 ms to elicit antidromic responses from ChR2^{+} neurons. Occasionally, slightly longer pulse durations were used

(3 or 5 ms). Antidromic responses were seen on one or two recording channels per recording session.

Two-photon calcium imaging. To label intratelencephalic cells for imaging, cholera toxin subunit B (CTB; Alexa 647; Molecular probe, Invitrogen, 0.5% in HEPES buffered saline) was injected to the contralateral (right) ALM (2.5 mm anterior, 1.5 mm lateral to bregma, 300 and 600 μm deep, 100 nl per site). For pyramidal tract cells, red RetroBeads (Lumafluor) were injected into the ipsilateral (left) basal pontine nucleus (3.5 mm posterior, 0.4 mm lateral, 5, 5.4, and 5.8 mm below brain surface, 100 nl per site). Window surgery and GCaMP virus injections were carried out 12–34 days after tracer injection. A circular craniotomy (~3 mm diameter) was made above left ALM (centred at 2.5 mm anterior and 1.5 mm lateral to bregma). AAV2/1-syn-GCaMP6s-WPRE virus (UPenn Viral Core, AV-1-PV2824) was diluted two- to sixfold in HEPES buffered saline. Injections were made at three to five locations centred around ALM (separated by ~400 μm) and at three depths (210/370/530 μm) for each location (~5–6 nl per depth). The imaging window was constructed from two layers of microscope coverglass[10] and fixed to the skull using cyanoacrylate glue and dental acrylic. A metal post for head fixation was implanted posterior to the window using dental acrylic. Water restriction started 5–7 days after window surgery. Behavioural training started ~5–7 days after water restriction.

Imaging experiments started after the animals had learned the task (>70% trials correct; typically ~20–30 days after surgery). Images were acquired using a custom-built two-photon microscope equipped with a resonant galvo scanning module (Thorlabs), controlled by ScanImage 4.2 (http://www.scanimage.org). The light source was a femtosecond pulsed laser (Coherent). The objective was a 16× water immersion lens (Nikon, 0.8 NA, 3 mm working distance). GCaMP6s was excited at 940 nm and images (512 × 512 pixels, 400 μm × 400 μm or 600 μm × 600 μm) were acquired at 15 Hz. The average excitation power was up to 120 mW for L5 neurons. After functional imaging of a particular *z* plane, the laser wavelength was switched to 830 nm to image Retrobead and CTB-647. A small image stack was acquired around the imaging location to allow unambiguous identification of cells that were out of focus. Retrobead and CTB-647 were imaged with Brightline 609/54 (Semrock) and HQ-675/70-2P (Chroma), respectively.

Behavioural data analyses. Performance was computed as the fraction of correct reports, excluding the 'lick early' trials. Chance performance was 50%. We also separately computed the performance for lick right and lick left trials (Figs 1 and 5). Behavioural effects of photoinhibition (Fig. 1) and photoactivation (Fig. 5) were quantified by comparing the performance under photostimulation with control performance (Figs 1 and 5). Significance of the performance change in each photostimulation condition was determined using bootstrapping to account for variability across mice, sessions and trials. We tested against the null hypothesis that the performance change caused by photostimulation was due to normal behavioural variability. In each round of bootstrapping, we replaced the original behavioural data set with a re-sampled data set in which we re-sampled with replacement from: (1) animals; (2) sessions performed by each animal; and (3) the trials within each session. We then computed the performance change on the re-sampled data set. Repeating this procedure 10^4 times produced a distribution of performance changes that reflected the behavioural variability. The *P* value of the observed performance change was computed as the fraction of times the bootstrapping produced an inconsistent performance change (for example, if a performance decrease was observed during photostimulation, the *P* value is the fraction of times a performance increase was observed during bootstrapping, one-tailed test).

Electrophysiology data analyses. The extracellular recording traces were band-pass filtered (300 Hz–6 kHz). Events that exceeded an amplitude threshold (four standard deviations of the background) were subjected to manual spike sorting to extract single units[13]. A total of 1,408 single units were recorded across 99 recording sessions. Most recorded single units were in layer 5a and 5b. Depths were inferred from manipulator depth and calibration experiments[13]. Spike widths were computed as the trough-to-peak interval in the mean spike waveform (Extended Data Fig. 1). Units with spike width <0.35 ms were defined as fast-spiking neurons (124/1,408) and units with spike width >0.45 ms as putative pyramidal neurons (1,245/1,408). This classification was previously verified by optogenetic tagging of GABAergic neurons[13]. Units with intermediate values (0.35–0.45 ms, 39/1,408) were excluded from our analyses. Fast-spiking neurons had higher baseline firing rates and were less selective than putative pyramidal neurons (Extended Data Fig. 1). We concentrated our analyses on the putative pyramidal neurons.

Neurons were further tested for significant trial-type selectivity during the sample, delay or response epochs, using the spike counts from the lick left and lick right trials (two-tailed *t*-test, $P < 0.05$; Fig. 2). Neurons that significantly differentiated trial types during any one of the trial epochs were deemed 'selective' (912/1,245). Neurons with selectivity during the sample or delay epochs were classified as having 'preparatory activity'. Neurons with significant selectivity during the response epoch were classified as having 'peri-movement selectivity' (Fig. 2c, d). Selective neurons were classified into 'contra-preferring' versus 'ipsi-preferring', based on their total spike counts across all three trial epochs (Fig. 2e). To compute selectivity (Fig 2c), we first determined each neuron's preferred trial type using spike counts from a subset of the trials (10 trials); selectivity is calculated as the spike rate difference between the trial types on the remaining data. We quantified the dynamics of the ALM population response by computing Pearson's correlation between population response vectors at different times and the population response vector at the onset of the go cue (Fig. 2c, bottom panel). We assumed all neurons were recorded simultaneously (ignoring potential correlations between neurons). To equalize the contributions of individual neurons, each neuron's response was mean-subtracted and normalized to the variance of its response across the entire trial epoch (computed in time bins of 200 ms). To compute the contra-selectivity (Fig. 3), we took the firing rate difference between the lick right trials and lick left trials for each neuron. These firing rate differences were averaged across the selective neurons. Only trials in which mice correctly reported pole locations were included. Standard errors of the mean were obtained by bootstrapping (Figs 2 and 3). Bootstrapping was also used to evaluate whether contra-preferring neurons were significantly higher in proportion (Fig. 3f, j). The neuronal data set was re-sampled with replacement, and the *P* value reflected the fraction of times when more ipsi-preferring neurons were observed (one-tailed test against the null hypothesis that there were more ipsi-preferring neurons).

Seventy-three neurons were antidromically activated by photostimulating pyramidal tract neuron axons, and 134 neurons were antidromically activated by photostimulating intratelencephalic neuron axons. These neurons were further tested for collisions in which we looked for absence of antidromic spikes when preceded by spontaneous spikes (Extended Data Fig. 5)[6]. Neurons that passed the collision test were classified as pyramidal tract neurons (45/73) or intratelencephalic neurons (27/134). Neurons that failed the test were classified as pyramidal-tract-activated (22/73) or intratelencephalic-activated (106/134) because they were presumably synaptically-connected to $ChR2^+$ neurons and intratelencephalic neurons. A few cells could not be tested due to an absence of spontaneous activity (6/73 from pyramidal tract neuron tagging; 1/134 from intratelencephalic neuron tagging) and these neurons were excluded from further analyses. Pyramidal tract neurons and intratelencephalic neurons had shorter mean antidromic spike latencies (pyramidal tract neurons, 2.7 ms; intratelencephalic neurons, 4.6 ms) and smaller temporal jitter (mean s.d. of the antidromic spike latency, pyramidal tract neurons, 0.1 ms; intratelencephalic neurons, 0.2 ms) than the pyramidal-tract-activated (mean latency, 8.3 ms; jitter, 2.8 ms) and intratelencephalic-activated neurons (mean latency, 9.8 ms; jitter, 2.4 ms). However, the latency distributions overlapped and a neuron could not be reliably inferred as being a $ChR2^+$ neuron based on its response latency alone (Extended Data Fig. 5e, see example in Extended Data Fig. 5b, d). The classification of intratelencephalic neurons and pyramidal tract neurons was further corroborated by the following observations: first, intratelencephalic neurons had longer spike latencies than pyramidal tract neurons, consistent with previous reports of antidromic electrical stimulation in primates and cats[6,31] and the fact that pyramidal tract neurons have myelinated axons and fast conduction velocities. Second, our observed latency values and the estimated conduction velocities (pyramidal tract, 6.5 m s^{-1}; intratelencephalic, 2.3 m s^{-1}) were in agreement with previous measurements from intratelencephalic neurons and pyramidal tract neurons in mice[50,51]. Finally, intratelencephalic neurons showed greater temporal jitter in response to photostimulation than pyramidal tract neurons (Fig. 3h, Extended Data Fig. 5). However, following a conditioning stimulus, the responses of intratelencephalic neurons to a second light pulse (10 ms later) showed less temporal jitter (Extended Data Fig. 5g). This effect has been described for corpus-callosum-projecting intratelencephalic neurons during electrical stimulation in cats[52].

Calcium imaging data analyses. The brain motion was corrected using a line-by-line correction algorithm[10]. Regions of interest (ROIs) corresponding to identifiable cell bodies were selected using a semi-automated algorithm[32]. Individual neurons were visually identified on average fluorescence images as well as a pixel-based response map and a 'neighborhood correlation map' (where the brightness of each pixel encodes the correlation of its fluorescent time course to that of its immediate neighbours) that highlight task-related and active cells, respectively[53]. The fluorescence time course of each cell was measured by averaging all pixels within the ROI, with a correction for neuropil contamination[54]. The fluorescence signal of a cell body was estimated as $F_{\mathrm{cell\_true}}(t) = F_{\mathrm{cell\_measured}}(t) - r \times F_{\mathrm{neuropil}}(t)$, with $r = 0.7$. The neuropil signal $F_{\mathrm{neuropil}}(t)$ surrounding each cell was measured by averaging the signal of all pixels within a ~20 μm region from the cell centre (excluding all selected cells). To ensure robust neuropil subtraction, only cells that were at least 5% brighter than the surrounding neuropil were included. $\Delta F/F_0$ was calculated as $(F-F_0)/F_0$, where F_0 is the baseline fluorescence signal averaged over a 0.5 s period immediately before the start of each trial.

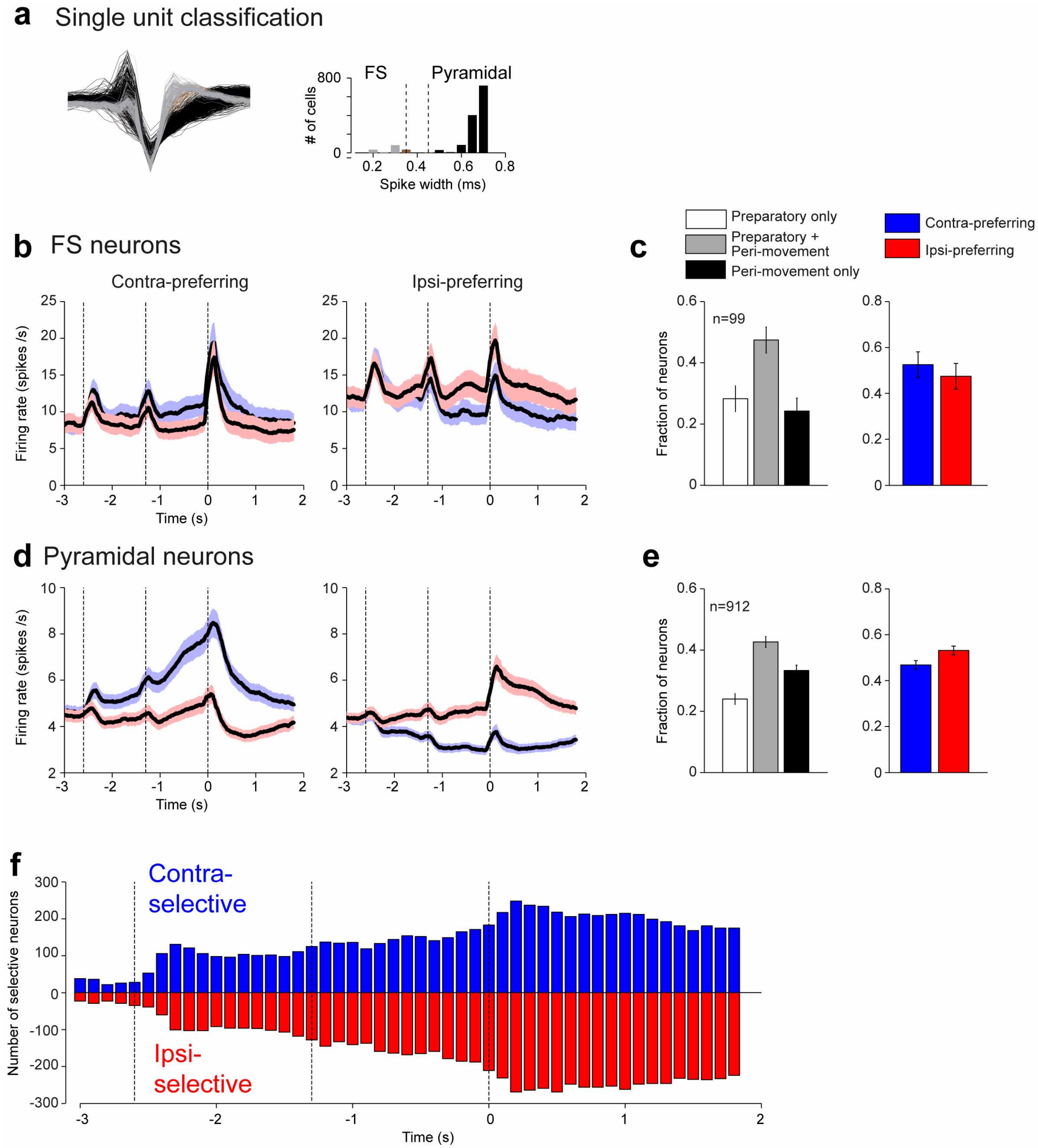
Task-related neurons were defined as neurons showing significant fluorescence modulation during the task. This was calculated using non-parametric ANOVA (Kruskal–Wallis test) across multiple 0.33 s time bins (five image samples) during

the task. A cell was classified as task-related if the null hypothesis that all time bins had equal fluorescence can be rejected at a P value of 0.01 during either contra or ipsi trials. This criterion identified 58% (2,740/4,706) cells as task related. We further defined trial-type-selective cells as a subset of task related neurons that showed significantly different $\Delta F/F$ response during contra and ipsi trials ($P < 0.05$; Wilcoxon rank sum test). 57% (1575/2740) of task related cell are trial-type selective.

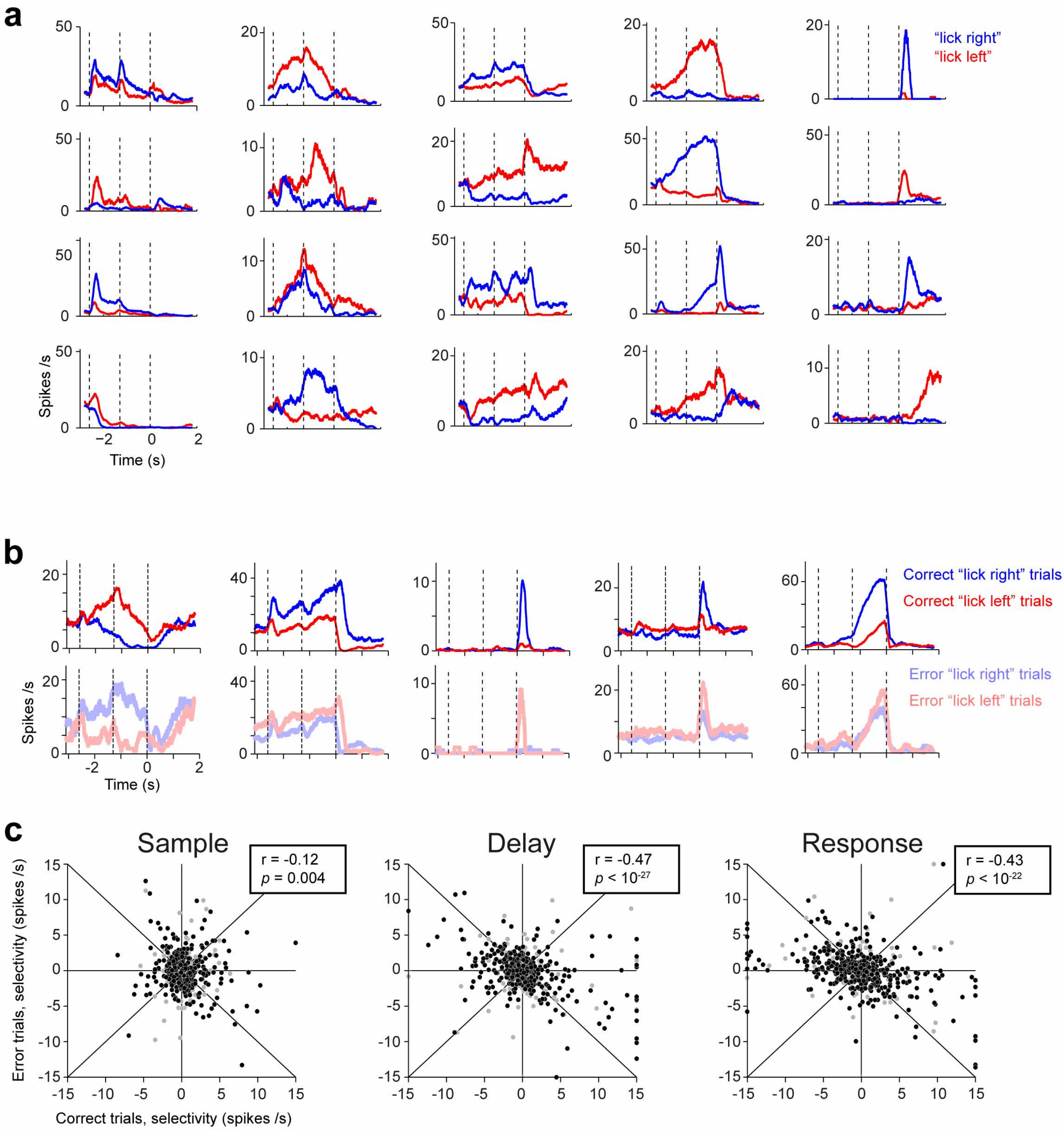
We examined spatial clustering of response types using two different measures. First, we quantified the probability that the nearest neighbour of each contra-preferring or ipsi-preferring cell was of the same response type, compared to the probability calculated by randomly shuffling the response type labels. The P value was the fraction of bootstrap trials with probability larger than the observed probability. Second, we quantify the distance between nearest neurons of the same response type and compare the distance to those calculated with randomized response type labels. The P value was the fraction of bootstrap trials with distance smaller than the observed distance.

Overall, imaging detected a slight majority of contra-preferring neurons (61%, 971/1,575), whereas no bias was detected based on spike counts (Fig. 2e). To resolve this discrepancy we used spiking data and a model derived from calibration experiments involving simultaneous *in vivo* imaging and recording[32] to simulate fluorescence dynamics (data not shown). The simulations revealed that imaging preferentially detects selectivity in neurons (1) with relatively strong spike rate differences between trial types; (2) responding with an increase, rather than a decrease, in spike rate. Overall this produced a bias for detecting contra-preferring neurons in the simulations, consistent with the observed bias in the imaging experiments. The comparison of simulation and imaging highlights that selectivity evolves differently in contra-preferring and ipsi-preferring neurons (see also Extended Data Fig. 1d). Contra-preferring neurons include a majority of the neurons with large spike count differences, whereas ipsi-preferring neurons tend to express their selectivity more often through contra-suppression.

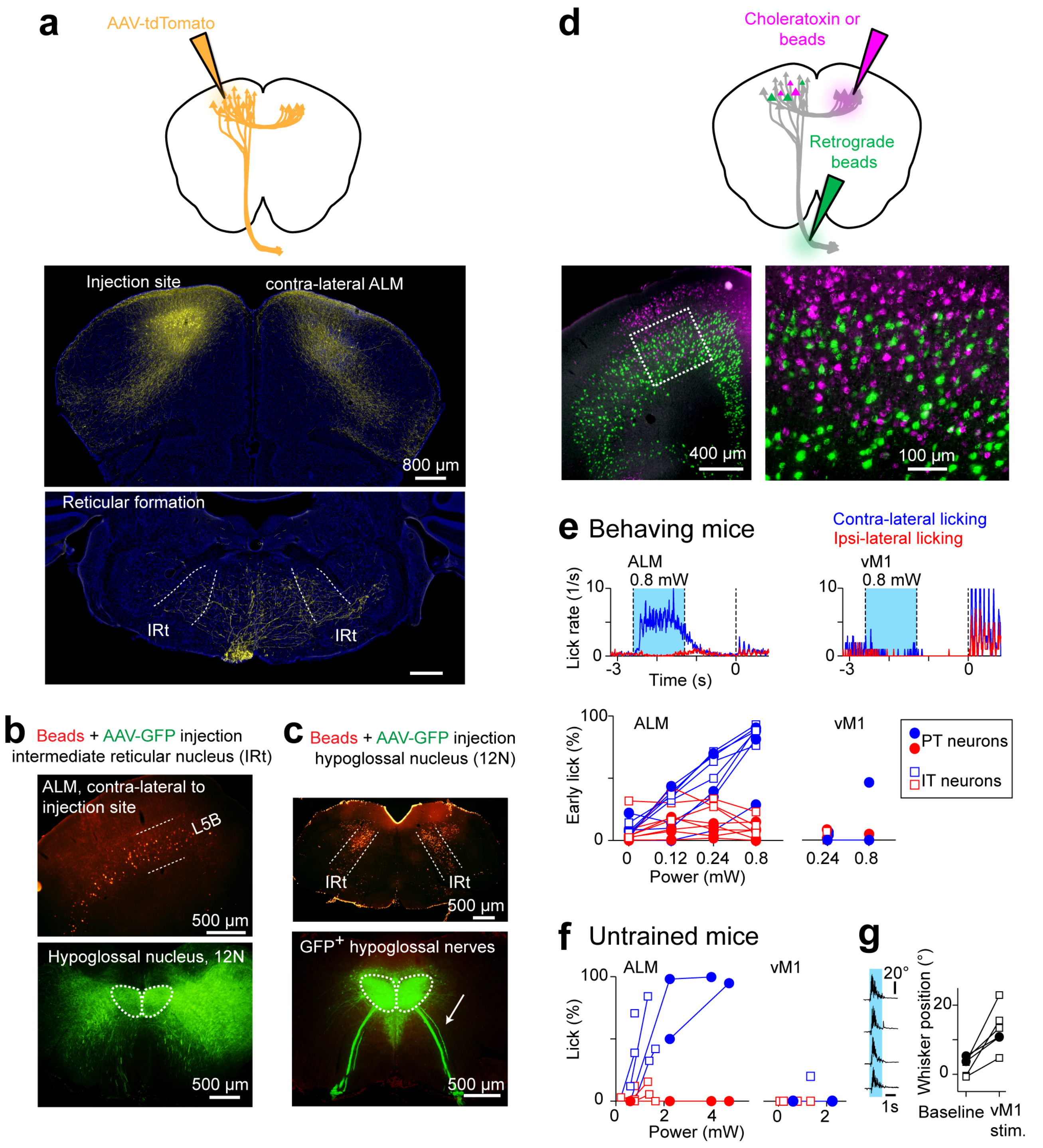
44. Zhao, S. *et al.* Cell type-specific channelrhodopsin-2 transgenic mice for optogenetic dissection of neural circuitry function. *Nature Methods* **8,** 745–752 (2011).
45. Hippenmeyer, S. *et al.* A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3,** e159 (2005).
46. Lin, J. Y., Knutsen, P. M., Muller, A., Kleinfeld, D. & Tsien, R. Y. ReaChR: a red-shifted variant of channelrhodopsin enables deep transcranial optogenetic excitation. *Nature Neurosci.* **16,** 1499–1508 (2013).
47. Gerfen, C. R., Paletzki, R. & Heintz, N. GENSAT BAC Cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. *Neuron* **80,** 1368–1383 (2013).
48. Madisen, L. *et al.* A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nature Neurosci.* **15,** 793–802 10.1038/nn.3078 (2012).
49. O'Connor, D. H. *et al.* Neural coding during active somatosensation revealed using illusory touch. *Nature Neurosci.* **16,** 958–965 (2013).
50. Simmons, P. A. & Pearlman, A. L. Receptive-field properties of transcallosal visual cortical neurons in the normal and reeler mouse. *J. Neurophysiol.* **50,** 838–848 (1983).
51. Alstermark, B. & Ogawa, J. In vivo recordings of bulbospinal excitation in adult mouse forelimb motoneurons. *J. Neurophysiol.* **92,** 1958–1962 (2004).
52. Swadlow, H. A., Waxman, S. G. & Rosene, D. L. Latency variability and the identification of antidromically activated neurons in mammalian brain. *Exp. Brain Res.* **32,** 439–443 (1978).
53. Chen, T.-W. *A Systems Level Analysis of Neuronal Network Function in the Olfactory Bulb: Coding, Connectivity, and Modular organization*, PhD Thesis. University of Göttingen, Göttingen, Germany (2008).
54. Kerlin, A. M., Andermann, M. L., Berezovskii, V. K. & Reid, R. C. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron* **67,** 858–871 (2010).



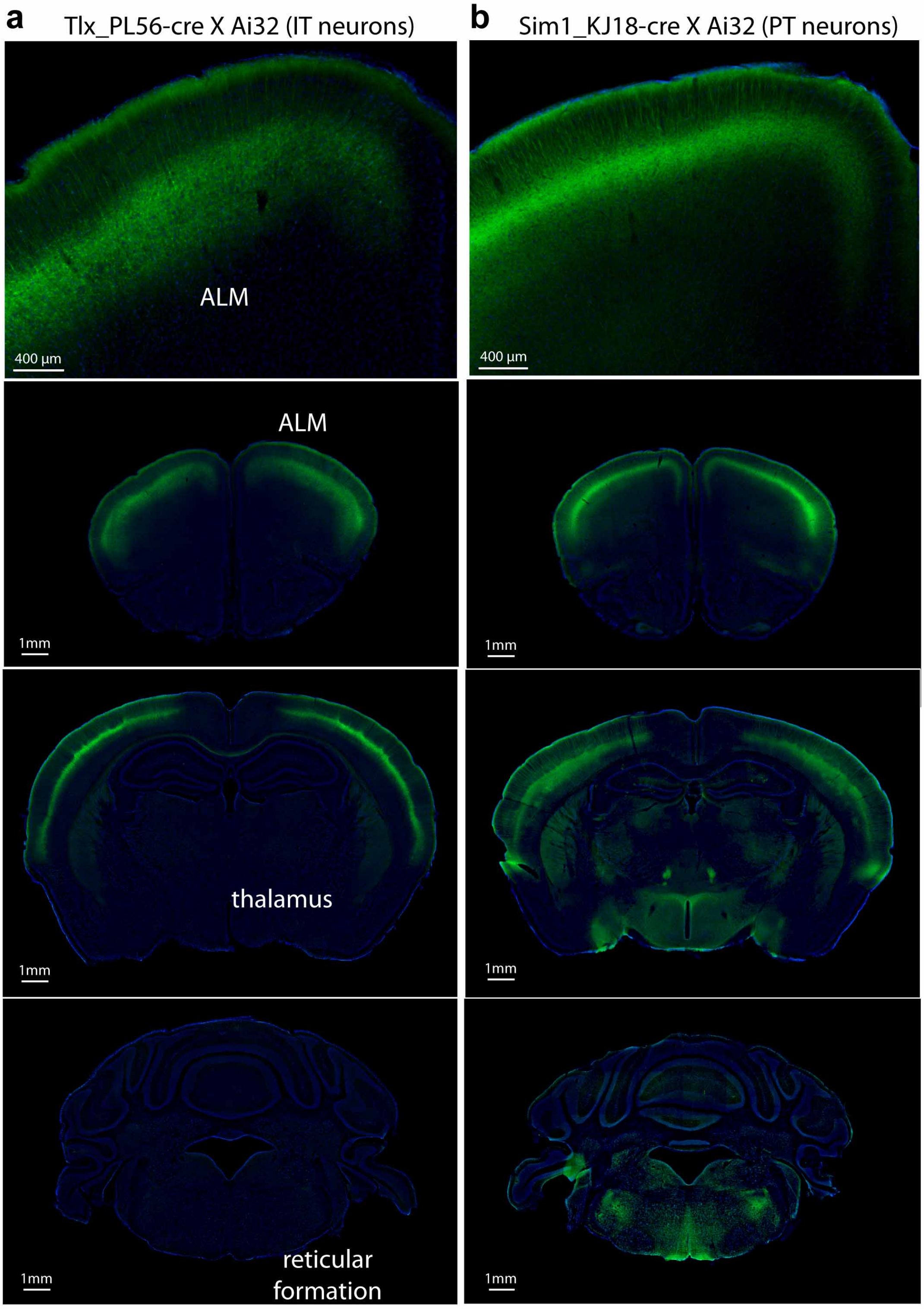
Extended Data Figure 1 | Neural selectivity in ALM. a, Single-unit classification. Left, overlaid mean spike waveforms for putative fast-spiking (FS) interneurons (grey, $n = 124$) and putative pyramidal neurons (black, $n = 1245$). A small subset of single units with intermediate spike durations were not classified (brown, $n = 39$). Right, histogram of spike durations (Methods). **b**, Fast-spiking neurons population response (mean ± s.e.m.) during the lick right trials (contralateral, blue) and lick left trials (ipsilateral, red). Neurons are sorted by their preferred trial type using spike counts from 10 trials and the remaining data was used to compute the selectivity. Left, contra-preferring neurons. Right, ipsi-preferring neurons. **c**, Left, proportion of neurons in **b** with preparatory and peri-movement activity. Right, contra-preferring versus ipsi-preferring selectivity. Error bars, s.e.m. across animals, bootstrap. **d**, **e**, Same as (**b**) and (**c**) but for putative pyramidal neurons. **f**, Number of significantly selective putative pyramidal neurons as a function of time. Significant selectivity was based on spike counts in 200-ms time windows, $P < 0.05$, two-tailed *t*-test. Dashed lines, behavioural epochs.



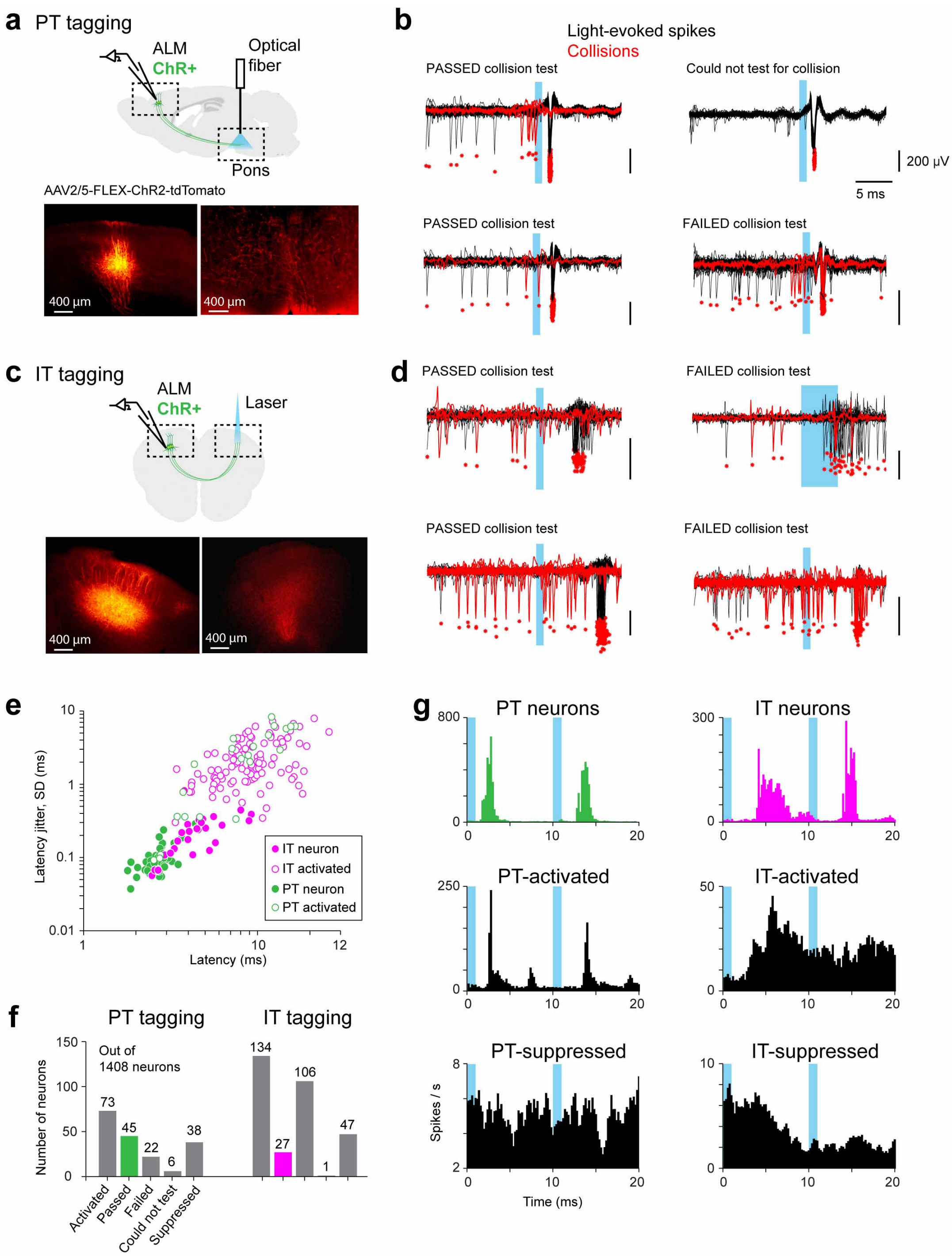
Extended Data Figure 2 | ALM neurons exhibit temporally complex responses and choice-specific selectivity. **a**, Twenty example ALM neurons responding during different epochs of the object location discrimination. Correct lick right (blue) and lick left (red) trials only. Dashed lines, behavioural epochs. Averaging window, 200 ms. **b**, Six example ALM neurons during object location discrimination. Top, peri-stimulus time histogram for correct lick right and lick left trials. Bottom, peri-stimulus time histogram for error trials (transparent colour). **c**, ALM neurons show choice-specific preparatory activity. Selectivity is the firing rate difference between lick right and lick left trials during sample, delay or response epochs ((firing rate lick right)–(firing rate lick left)). Circles, individual neurons ($n = 912$). Filled circles, neurons with significant selectivity ($P < 0.05$, two-tailed t-test). On error trials, when mice licked in the opposite direction to the instruction provided by object location (Fig. 1a), a majority of ALM neurons switched their trial type preference to predict the licking direction, as indicated by the negative correlations (r, Pearson's correlation).



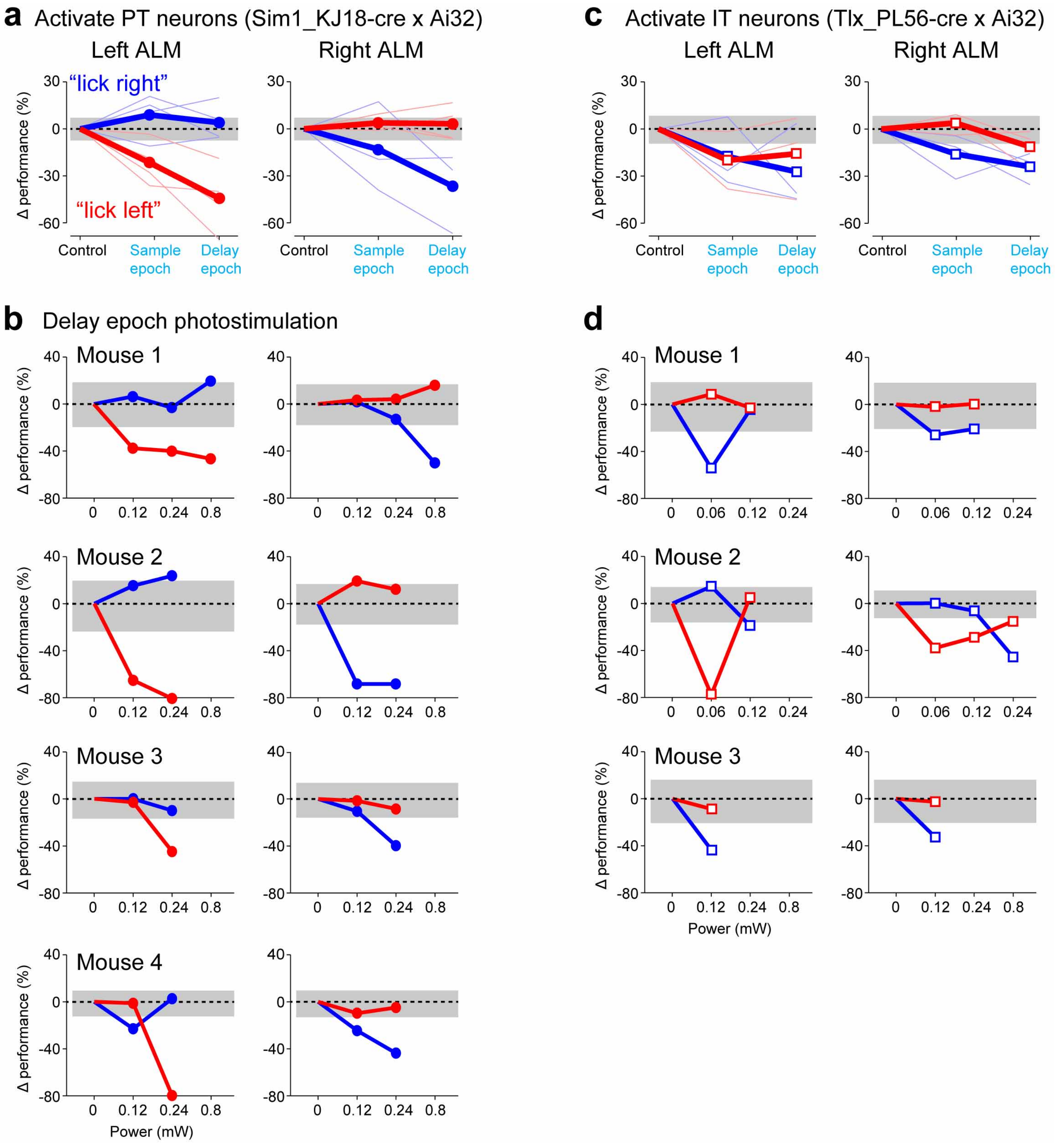
Extended Data Figure 3 | ALM pyramidal tract neurons control contralateral licking. **a**, Axonal projections of intratelencephalic neurons (top, Tlx_PL56 mice) and pyramidal tract neurons (bottom, Sim1_KJ18 mice). **b**, Top, retrogradely labelled pyramidal tract neurons in contralateral ALM. Bottom, intermediate nucleus of the reticular formation (IRt) axonal projections in the hypoglossal nucleus, 12N. **c**, Top, retrogradely labelled IRt neurons. Bottom, hypoglossal nerves. **d**, Retrograde labelling of pyramidal tract neurons (green) and intratelencephalic neurons (magenta). The ALM coronal slice shows intermingled labelling of pyramidal tract neurons and intratelencephalic neurons without overlap. **e**, Unilateral stimulation of ALM pyramidal tract or intratelencephalic neurons triggered contralateral licking during behaviour (blue, contralateral licking; red, ipsilateral licking). Top, average lick rate during ALM (left) and left vibrissal motor cortex (vM1) (right) photostimulation, $n = 8$ mice. Dashed lines, behavioural epochs. Cyan region, photostimulation. Bottom, fraction of trials in which photostimulation caused 'early lick' as a function of laser power. Sample and delay epoch photostimulation data were combined. IT, intratelencephalic; PT, pyramidal tract. **f**, Unilateral stimulation of ALM pyramidal tract or intratelencephalic neurons triggered contralateral licking in untrained mice ($n = 6$ mice). Fraction of trials in which photostimulation caused licking as a function of laser power. **g**, Unilateral stimulation of left vM1 pyramidal tract or intratelencephalic neurons triggered whisker movements in untrained mice. Left, whisker azimuthal angle traces, individual trials. Right, average whisker angle 500 ms before photostimulation (baseline) and during vM1 photostimulation ($n = 6$ mice).



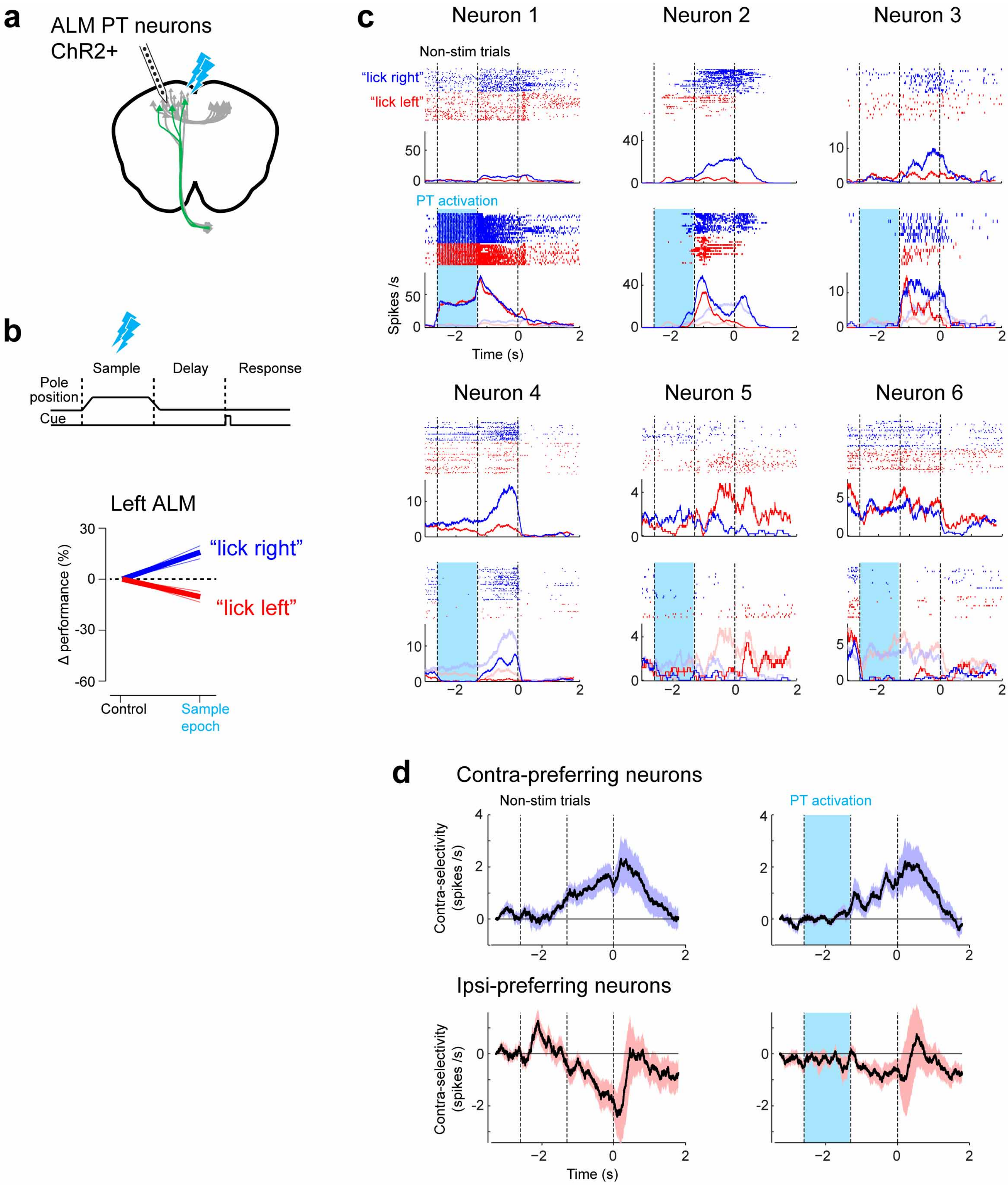
Extended Data Figure 4 | Transgenic ChR2 expression in intratelencephalic and pyramidal tract neurons. **a**, ChR2 expression in layer 5 intratelencephalic neurons. Tlx_PL56-Cre mouse crossed to a Rosa-ChR2-eYFP reporter mouse (Ai32). Top, ChR2 expression in ALM. Bottom, ChR2 expression in three coronal sections from anterior to posterior. **b**, Same as **a** for ChR2 expression in pyramidal tract neurons. Sim1_KJ18-Cre mouse crossed to Ai32.



Extended Data Figure 5 | Cell-type-specific recording with ChR2 tagging. **a**, Pyramidal tract neuron recordings. Top, left ALM neurons were infected with AAV2/5-FLEX-ChR2-tdTomato (Sim1_KJ18-Cre mice). An optical fibre was implanted into the left (ipsilateral) reticular formation to antidromically stimulate the pyramidal tract neuron axons. Pyramidal tract neurons were identified based on back-propagating antidromic spikes (tagging). Bottom, expression of ChR2-tdTomato in left ALM; injection site (left), axon terminals in the reticular formation (right). **b**, Recording traces for four example neurons activated by antidromic stimulation of pyramidal tract neurons (blue). Red ticks, individual spikes. Red traces, collision tests. Left, two pyramidal tract neurons that passed the collision test; antidromic spikes were absent when preceded by spontaneous spikes. These neurons were classified as pyramidal tract neurons. Right, two neurons that failed the collision test. The top neuron could not be tested for collision due to an absence of baseline firing. The bottom neuron failed the collision test because light-evoked spikes occurred even when preceded by spontaneous spikes. This neuron was classified as a pyramidal-tract-activated neuron. Scale bars, 200 µV, 5 ms. **c**, **d**, Same as **a** and **b** but for intratelencephalic neuron recordings. **e**, Antidromic spike latency versus jitter (SD) for pyramidal tract neurons, intratelencephalic neurons, pyramidal-tract-activated neurons, and intratelencephalic-activated neurons. **f**, Recording yield. 'Activated', neurons activated by antidromic stimulation; 'Passed', subsets of activated neurons that passed the collision test. These neurons were classified as pyramidal tract neurons or intratelencephalic neurons. 'Failed', subsets of activated neurons in which spontaneous spikes failed to block light-evoked spikes. 'Could not test', the subset of activated neurons without spontaneous activity. These neurons were excluded from analyses. 'Suppressed', neurons suppressed by antidromic stimulation. **g**, Top, light-evoked responses across all pyramidal tract neurons and intratelencephalic neurons. Middle, pyramidal-tract-activated neurons and intratelencephalic-activated neurons. Bottom, pyramidal-tract-suppressed neurons and intratelencephalic-suppressed neurons.



Extended Data Figure 6 | Photostimulation of intratelencephalic and pyramidal tract neurons biases upcoming licking direction. **a**, Performance change, pyramidal tract neurons activation. Re-plot of the data in Fig. 5b. Grey areas represent 95% confidence interval of expected behavioural variability. We estimated the behavioural variability by computing performance changes on re-sampled data sets in which we sampled with replacement from only the control trials (Methods, repeated 10^4 times). The number of trials in the re-sampled data sets was matched to the actual experiments. **b**, Dose response for individual mouse in **a**. Delay epoch photostimulation data only. **c**, **d**, Same as **a** and **b** but for intratelencephalic neurons activation.



Extended Data Figure 7 | Photostimulation of pyramidal tract neurons during the sample epoch causes persistent changes in ALM activity and a directional bias. **a**, Simultaneous recordings and photostimulation of left ALM pyramidal tract neurons during behaviour. Two mice, eight sessions, 91 neurons. **b**, Stimulation of left ALM pyramidal tract neurons during the sample epoch biased the upcoming licking to the contralateral direction. Top, sample epoch photostimulation. Bottom, performance change relative to the control trials. Thick lines, mean; thin lines, individual mice. **c**, Spike raster plots and peri-stimulus time histograms of six example neurons. Correct lick right (blue) and lick left (red) trials only. Dashed lines, behavioural epochs. For each neuron: top, control trials; bottom, photostimulation trials (solid colour). Control trials are overlaid in transparent colour. Averaging window, 200 ms. **d**, Average population selectivity (black line, ±s.e.m. across neurons). Left, control trials. Right, photostimulation trials. Neurons are sorted by their preferred trial type (top, contra-preferring neurons; bottom, ipsi-preferring neurons). Selectivity is the difference in spike rate between the preferred and non-preferred trial type.

doi:10.1038/nature14228

A gp130–Src–YAP module links inflammation to epithelial regeneration

Koji Taniguchi[1,2,3,4], Li-Wha Wu[1,5], Sergei I. Grivennikov[1,6], Petrus R. de Jong[7], Ian Lian[2,8,9], Fa-Xing Yu[2,8,10], Kepeng Wang[1], Samuel B. Ho[11], Brigid S. Boland[12], John T. Chang[12], William J. Sandborn[12], Gary Hardiman[13,14], Eyal Raz[7], Yoshihiko Maehara[3], Akihiko Yoshimura[4,15], Jessica Zucman-Rossi[16,17], Kun-Liang Guan[2,8] & Michael Karin[1,2,8]

Inflammation promotes regeneration of injured tissues through poorly understood mechanisms, some of which involve interleukin (IL)-6 family members, the expression of which is elevated in many diseases including inflammatory bowel diseases and colorectal cancer. Here we show in mice and human cells that gp130, a co-receptor for IL-6 cytokines, triggers activation of YAP and Notch, transcriptional regulators that control tissue growth and regeneration, independently of the gp130 effector STAT3. Through YAP and Notch, intestinal gp130 signalling stimulates epithelial cell proliferation, causes aberrant differentiation and confers resistance to mucosal erosion. gp130 associates with the related tyrosine kinases Src and Yes, which are activated on receptor engagement to phosphorylate YAP and induce its stabilization and nuclear translocation. This signalling module is strongly activated upon mucosal injury to promote healing and maintain barrier function.

Inflammation is a complex biological response triggered by tissue damage or microbial invasion. In addition to host defence, self-limiting inflammation triggers regeneration and repair[1,2]. By preventing further microbial translocation, healing promotes resolution of inflammation. Whereas host defence and immunity have been extensively studied, the mechanisms by which inflammation stimulates regenerative responses remain obscure. Numerous pathways involved in tissue growth, patterning and differentiation are re-deployed during regeneration, including the Hedgehog (Hh)–Gli, Wnt–β-catenin, Notch and Hippo–YAP pathways[3,4]. Upon tissue injury, myeloid cells, including macrophages, produce inflammatory cytokines and growth factors[5]. However, signalling mechanisms that link typical inflammatory cytokines to pivotal transcriptional regulators of tissue growth, repair and regeneration remain to be charted.

Regenerative responses are particularly important in the mammalian gastrointestinal tract, a tissue subject to frequent erosion and renewal. Unrepaired mucosal injury disrupts the epithelial barrier that prevents translocation of intestinal microbiota, resulting in acute inflammation[6]. Persistent failure to repair such damage can result in inflammatory bowel diseases (IBD), including ulcerative colitis, which entails severe mucosal erosion, and Crohn's disease (CD), in which aberrant growth can cause fistula formation[6]. Mucosal healing is a key treatment goal in IBD that predicts sustained remission and resection-free survival[6]. It is therefore important to understand how mucosal healing is regulated. After injury, intestinal epithelial cells (IECs) surrounding the lesion lose columnar polarity and rapidly initiate wound healing. 'Epithelial restitution' starts within minutes of injury and is considered crucial for temporary sealing of the disrupted surface. Subsequent stem-cell activation, proliferation and differentiation increase the cell pool available for healing. These processes are tightly regulated to prevent uncontrolled proliferation and tumorigenesis, and rely on coordinated and balanced function of IECs, secretory cells, intestinal stem cells and the immune system[6].

IL-6 is a prototypical pro-inflammatory cytokine, with family members including IL-11, IL-27, IL-31, leukaemia inhibitory factor (LIF), oncostatin M, ciliary neurotrophic factor, and cardiotrophin-1, all of which influence cell proliferation, survival, migration, invasion, angiogenesis and inflammation[7]. Most family members activate the JAK–STAT3, SHP2–Ras–ERK and PI(3)K–AKT–mTORC1 pathways via the common co-receptor gp130 (refs 7, 8). Among these pathways, STAT3 is the major and most extensively studied effector that links inflammation to cell proliferation, survival and cancer, being subject to feedback regulation by suppressor of cytokine signalling 3 (SOCS3)[8,9]. IL-6, soluble IL-6Rα and IL-11 are highly elevated in IBD and gastrointestinal cancers[10,11]. However, activating STAT3 mutations are rare[12], and tumoral STAT3 is mainly activated by cytokines and/or decreased SOCS3 expression[13]. Nonetheless, gain-of-function mutations affecting gp130–STAT3 signalling were identified in benign human inflammatory hepatocellular adenomas (IHCA)[12,14].

IL-6 promotes IEC proliferation and regeneration, and IL-6-deficient mice, which do not exhibit developmental abnormalities, are highly sensitive to experimental colitis induced by dextran sulfate sodium salt (DSS)[13]. Correspondingly, IL-6 blockade in humans can result in intestinal perforation[15]. In addition to STAT3 in IECs[13], mucosal regeneration after DSS challenge requires concomitant activation of Yes-associated protein (YAP)[16] and Notch[17]. YAP is a key transcriptional co-activator of tissue growth, which is normally kept inactive in the cytoplasm through serine phosphorylation by the Hippo effector kinase LATS[18]. YAP is activated either upon inhibition of Hippo signalling or upon tyrosine phosphorylation by the Src family kinase (SFK) Yes[19]. Notch is activated

[1]Laboratory of Gene Regulation and Signal Transduction, University of California, San Diego, La Jolla, California 92093, USA. [2]Departments of Pharmacology and Pathology, University of California, San Diego, La Jolla, California 92093, USA. [3]Department of Surgery and Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka 812-8582, Japan. [4]Department of Microbiology and Immunology, Keio University School of Medicine, Tokyo 160-8582, Japan. [5]Institute of Molecular Medicine, College of Medicine, National Cheng Kung University, Tainan 70101, Taiwan. [6]Fox Chase Cancer Center, Cancer Prevention and Control Program, Philadelphia, Pennsylvania 19111, USA. [7]Department of Medicine, University of California, San Diego, La Jolla, California 92093, USA. [8]Moores Cancer Center, University of California, San Diego, La Jolla, California 92093, USA. [9]Department of Biology, Lamar University, PO Box 10037, Beaumont, Texas 77710, USA. [10]Children's Hospital and Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China. [11]Department of Medicine, VA San Diego Healthcare System, San Diego, California 92161, USA. [12]Inflammatory Bowel Disease Center, Division of Gastroenterology, Department of Medicine, School of Medicine, University of California, San Diego, La Jolla, California 92093, USA. [13]Department of Medicine, Medical University of South Carolina, Charleston, South Carolina 29425, USA. [14]CSRC and BIMRC, San Diego State University, San Diego, California 92182, USA. [15]Japan Science and Technology Agency, CREST, Tokyo 102-0076, Japan. [16]Inserm, UMR 1162, Génomique fonctionnelle des tumeurs solides, IUH, Paris 75010, France. [17]Université Paris Descartes, Labex Immuno-oncology, Sorbonne Paris Cité, Faculté de Medicine, Paris 75006, France.

by ligands such as Jagged 1, Jagged 2 and Delta-like (DLL) 1, 3 and 4, which trigger Notch cleavage by γ-secretase, resulting in nuclear translocation of its intracellular domain (NICD) which associates with CBF1/RBP-Jκ to activate target gene transcription[20]. The mechanisms whereby mucosal injury activates YAP and Notch remain elusive.

We show that independently of STAT3, gp130 also activates YAP and then Notch through direct association with SFKs. This pathway is engaged upon mucosal injury in mice and is essential for inflammation-induced epithelial regeneration; it is also activated in human IBD.

gp130 activation causes aberrant IEC proliferation

We generated *villin-gp130*^{Act} transgenic mice that express activated gp130^{Act} (an in-frame S187–Y190 deletion found in IHCA[14]) from the IEC-specific villin promoter (Extended Data Fig. 1a, b). Hemizygous *villin-gp130*^{Act} mice were born in Mendelian ratios, but their intestines were larger and longer than wild-type counterparts (Fig. 1a and Extended Data Fig. 1c). The transgenic intestinal mucosa was hyperproliferative with deeper crypts than wild type (Fig. 1b–d and Extended Data Fig. 1d), without a difference in apoptosis (Extended Data Fig. 1e). The villus-crypt structure was disorganized due to increased cell proliferation, but secretory cells (goblet, Paneth and enteroendocrine cells) were markedly decreased in the *villin-gp130*^{Act} small intestine (Fig. 1e, f and Extended Data Fig. 1f–h). Ectopic alkaline phosphatase (AP) staining in *villin-gp130*^{Act} small intestinal crypts suggested abortive differentiation of absorptive enterocytes (Fig. 1g). Electron microscopy revealed short, thick and non-uniform microvilli, typical of undifferentiated brush border cells (Extended Data Fig. 1i). Lamina propria CD45$^+$ immune cells were increased in the *villin-gp130*^{Act} intestine, and pro-inflammatory cytokines, such as *Il6* and *Tnf*, were upregulated (Extended Data Fig. 2a, b). However, more modest differences in the colonic mucosa were found

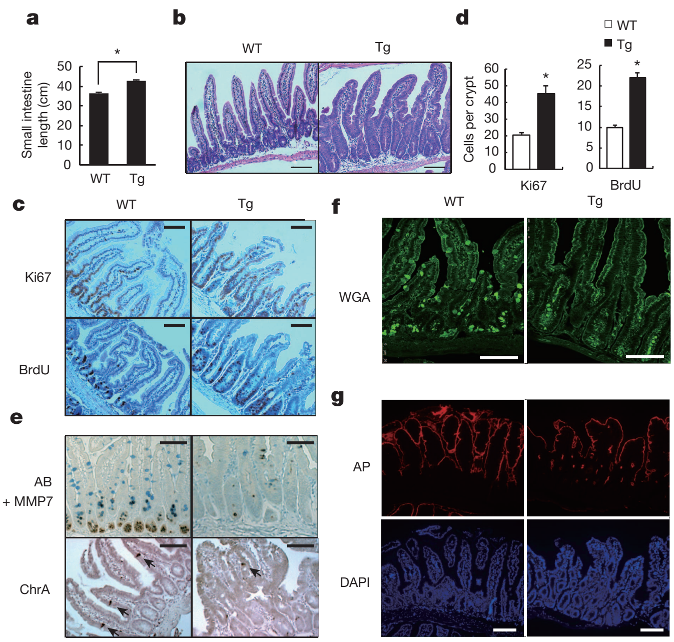


Figure 1 | Persistent gp130 activation causes aberrant IEC proliferation and differentiation. **a**, Wild-type (WT) and *villin-gp130*^{Act} (Tg) small intestine lengths at 3 months ($n = 5$). Data represent averages ± standard error of the mean (s.e.m.); $*P < 0.05$. **b**, Haematoxylin and eosin (H&E) staining of paraffin-embedded small intestinal sections from wild-type and *villin-gp130*^{Act} mice. Shown are representative images. **c**, **e**–**g**, Immunohistochemical analysis of paraffin-embedded small intestinal sections from wild-type and *villin-gp130*^{Act} mice ($n = 6$). Ki67 and BrdU incorporation (**c**), Alcian blue (AB) plus MMP7 and chromogranin A (ChrA) (**e**), wheat germ agglutinin (WGA) (**f**), and alkaline phosphatase (AP) and DAPI (**g**) stainings are shown. **d**, Ki67- and BrdU-positive cells were counted in each crypt. Data are averages ± s.d.; $*P < 0.05$. Scale bars represent 100 μm (**b**, **c**, **e**–**g**). Arrows in **e** indicate ChrA-positive cells.

between wild-type and transgenic mice (Extended Data Fig. 2c), probably because the villin promoter is more active in the small intestine[21].

gp130 activates YAP and Notch signalling

The gp130 effectors STAT3, STAT1 and ERK1/2 were activated in *villin-gp130*^{Act} mice (Fig. 2a, b and Extended Data Fig. 2d, e). Ribosomal protein S6 phosphorylation, indicating mTORC1 activation, was also elevated in *villin-gp130*^{Act} small intestine (Extended Data Fig. 2d). An increase in undifferentiated, proliferating enterocytes and a decrease in secretory cell lineages are typical of mice with strongly activated YAP or Notch[22–24], whereas Notch/γ-secretase inhibition converts proliferative IECs into goblet cells[25]. YAP and its target, connective tissue growth factor (CTGF), and the Notch target HES1 were strongly elevated in *villin-gp130*^{Act} small intestine, along with higher amounts of activated β-catenin (Fig. 2a). YAP was mainly found in IEC nuclei in villi and crypts (Fig. 2b), as well as in colon (Extended Data Fig. 2e). Nuclear HES1 was also elevated in *villin-gp130*^{Act} small intestinal crypts (Fig. 2b) and strong nuclear YAP staining was observed in *villin-gp130*^{Act} intestinal organoids (Fig. 2c), a useful *ex vivo* system for studying IEC signalling[26]. *Yap* mRNA expression remained unchanged (Extended Data Fig. 3a), suggesting post-transcriptional regulation. *Ctgf* and *Hes1*, however, were upregulated at the mRNA level along with mRNAs for intestinal stem-cell markers, *Cd44* and *Cd133*, and Notch ligands and receptors (Extended Data Fig. 3a, b). Transgenic intestinal organoids at day 3 after passage formed more rounded structures, although no morphological difference was seen between them and wild-type organoids on day 6 (Extended Data Fig. 3c). Rounded organoids form upon β-catenin activation[27], suggesting more active Wnt signalling in *villin-gp130*^{Act} organoids. YAP and Notch were also activated in human colorectal cancer (CRC) cells expressing a gp130 superactive variant (gp130^{SA}), which is refractory to feedback inhibition by SOCS3 due to an additional Y759F substitution (Extended Data Figs 1a and 3d). CD44 cleavage, another γ-secretase-mediated event[28], was upregulated in transgenic small intestine and

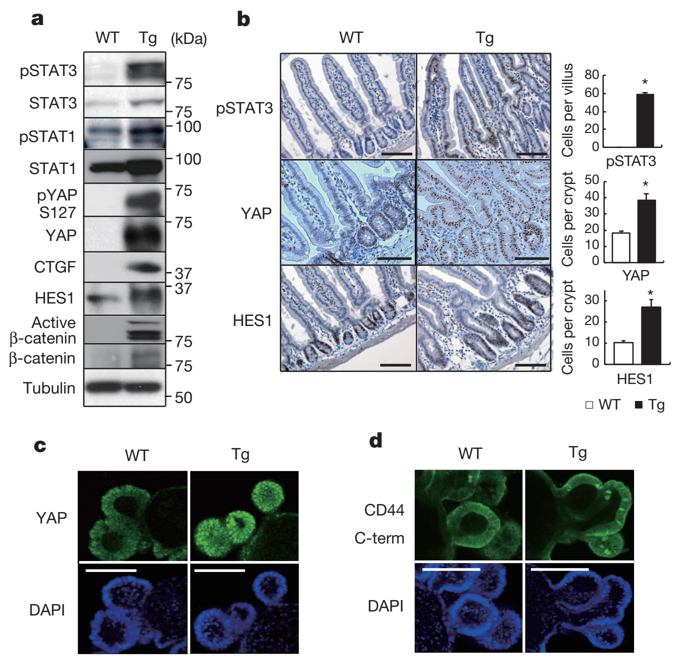


Figure 2 | gp130 activates YAP and Notch signalling. **a**, Lysates of wild-type and *villin-gp130*^{Act} jejuna were analysed for expression and phosphorylation of the indicated proteins. **b**, pSTAT3, YAP and HES1 stainings of paraffin-embedded small intestinal sections ($n = 6$). Positive cells were enumerated in each villus or crypt. Data are averages ± s.d.; $*P < 0.05$. **c**, **d**, Wild-type and *villin-gp130*^{Act} small intestinal organoids stained with YAP (**c**) and CD44 carboxy-terminal (**d**) antibodies were examined by immunofluorescent microscopy. Scale bars represent 100 μm (**b-d**) and all data are representative of at least 2–3 independent experiments.

organoids (Fig. 2d and Extended Data Fig. 2d). IL-6 activated YAP and STAT3 in human CRC cells and mouse primary hepatocytes (Extended Data Figs 3e–g).

Transcriptomic analysis confirmed the above results by showing strong enrichment for genes associated with innate immune response, response to viral infection and defence response in transgenic crypts (Extended Data Fig. 3h). Similar observations were made in IHCA patients with *gp130* activating mutations[14]. The growth factor amphiregulin (*Areg*), another YAP target[29], was upregulated in both *villin-gp130^{Act}* intestinal crypts and organoids (Extended Data Fig. 3h, i).

Notch or YAP inhibition restores IEC homeostasis

We examined whether Notch or YAP inhibition restored epithelial homeostasis in transgenic mice. To block Notch (and CD44) signalling we injected mice with a γ-secretase inhibitor, dibenzazepine (DBZ), and found that it restored secretory cell lineages, blocked HES1 expression and partially suppressed hyperproliferation and YAP activation (Fig. 3a, b and Extended Data Fig. 4a). To delete YAP in IECs, we crossed *villin-gp130^{Act}* mice with *villin-cre* × *Yap$^{fl/fl}$* (*Yap$^{\Delta IEC}$*) mice. YAP ablation largely restored secretory cell lineages and suppressed hyperproliferation and HES1 expression (Fig. 3c, d and Extended Data Fig. 4b). TAZ, a paralogue that partially compensates for YAP loss[30], was upregulated in the absence of YAP (Fig. 3e), possibly explaining the incomplete

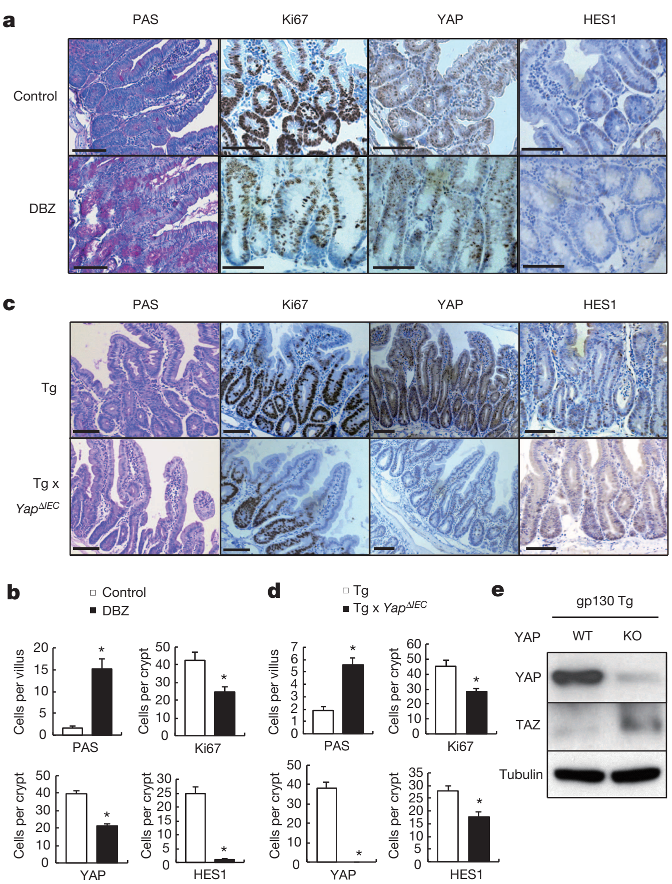


Figure 3 | Notch or YAP inhibition partially restores tissue homeostasis. **a**, **b**, PAS, Ki67, YAP and HES1 stainings of paraffin-embedded small intestinal sections from control and DBZ (10 μmol kg^{-1})-treated *villin-gp130^{Act}* (Tg) mice ($n = 3$ per group). Positive cells were enumerated in each villus or crypt (**b**). Data are averages ± s.d.; $*P < 0.05$. **c**, **d**, Paraffin-embedded small intestinal sections from *villin-gp130^{Act}* and *villin-gp130^{Act}/Yap$^{\Delta IEC}$* mice ($n = 4$ per group) were stained and quantified as above (**d**). Data are averages ± s.d.; $*P < 0.05$. **e**, Lysates of *villin-gp130^{Act}* and *villin-gp130^{Act}/Yap$^{\Delta IEC}$* jejuna were analysed for the indicated proteins. Scale bars represent 100 μm (**a**, **c**) and all data are representative of at least 2–3 independent experiments.

reversal of the *villin-gp130^{Act}* phenotype. We also investigated the contribution of other well-established gp130 effector pathways. We crossed *villin-gp130^{Act}* and *villin-cre* × *Stat3$^{fl/fl}$* (*Stat3$^{\Delta IEC}$*) mice. Notably, STAT3 ablation did not affect the *villin-gp130^{Act}* phenotype (Extended Data Fig. 4c). MEK (PD0325901) or PI(3)K (LY294002) inhibitors also did not reverse the *villin-gp130^{Act}* phenotype, although they both inhibited their targets (Extended Data Fig. 5a, b).

YAP upregulation by gp130-mediated SFK activation

We did not detect a decrease in YAP or LATS serine phosphorylation that could explain YAP activation in *villin-gp130^{Act}* IECs. On the contrary, both LATS and YAP S127 phosphorylation were elevated, suggesting enhanced Hippo activity (Fig. 2a and Extended Data Fig. 6a). Since Y357 phosphorylation by Yes activates YAP[19], we checked whether SFKs were activated in the transgenic small intestine. Human SFKs are positively regulated by Y419 phosphorylation and inhibited by Y530 phosphorylation[31]. Src/Yes Y419 was hyper-phosphorylated and Y530 was under-phosphorylated in *villin-gp130^{Act}* IECs (Fig. 4a). SFK activation was also observed in *villin-gp130^{Act}* intestinal organoids (Fig. 4b), human colon cancer cells overexpressing gp130^{Act} or stimulated with IL-6, correlating with YAP Y357 phosphorylation (Extended Data Figs 3e and 6b). YAP Y357 was also phosphorylated upon gp130^{Act} expression in intestinal organoids or human colon cancer cells (Fig. 4b and Extended Data Fig. 6c, d), after IL-11 stimulation (Extended Data Fig. 6e), or in primary hepatocytes stimulated with IL-6 and mouse liver undergoing partial hepatectomy, which also exhibited Notch and SFK activation (Extended Data Fig. 3g, 6f). YAP and Notch activation in liver were inhibited by treatment with the Src inhibitor PP2, indicating SFK dependence (Extended Data Fig. 6f). Y357 phosphorylation increases YAP protein stability[32]. Indeed, Src, but not JAK, inhibition accelerated YAP degradation in cycloheximide-treated cells (Extended Data Fig. 6g). Src inhibition also decreased YAP in organoids (Extended Data Fig. 7d), whereas Src activation enhanced YAP expression (Extended Data Fig. 6h). Notably, elevated Src Y419 phosphorylation and YAP expression were detected in approximately 60% of colonic biopsies collected from CD patients (Fig. 4d and Extended Data Table 1). Both SFK phosphorylation and YAP upregulation correlated with active disease.

Notably, gp130 interacted with endogenous Src and Yes, and both gp130 and YAP co-immunoprecipitated with Src (Fig. 4e and Extended Data Fig. 6i). Deletion of gp130 amino acids 812–827, which contain a phosphotyrosine motif, reduced binding to Src and attenuated activation of Src, YAP and Notch, but not STAT3 (Fig. 4f and Extended Data Fig. 6j, k). These data suggest that gp130 activates YAP through its interaction with SFK, independently of STAT3. Treatment of *villin-gp130^{Act}* mice with PP2 restored secretory cell lineages and inhibited IEC hyperproliferation, Src Y419 phosphorylation, HES1 expression, as well as ERK and YAP activation with only a modest effect on STAT3 (Fig. 4c and Extended Data Fig. 7a). PP2 and another SFK inhibitor (AZD0530), but not a JAK inhibitor (AZD1480) or DBZ, suppressed nuclear YAP in *villin-gp130^{Act}* intestinal organoids (Fig. 4g and Extended Data Fig. 7b, c). Consistent with a recent report that YAP potentiates β-catenin activation[33], *villin-gp130^{Act}* organoids contained more activated β-catenin than wild-type organoids, and Src inhibition reversed this effect (Extended Data Fig. 7d). Src, but not JAK inhibitors, blocked IL-6- or gp130^{Act}-induced YAP activation and stabilization in human CRC cells and mouse small intestinal organoids (Extended Data Fig. 7d–f).

gp130–SFK–YAP signalling aids mucosal regeneration

To examine gp130–SFK–YAP signalling during intestinal regeneration, we induced mucosal erosion with DSS. *villin-gp130^{Act}* mice exhibited less severe colitis and weight loss than wild-type mice, as well as reduced colon shortening and improved crypt architecture (Fig. 5a and Extended Data Fig. 8a, b), which are the opposite phenotypes of IEC-specific STAT3 or YAP deficiencies[13,16]. The colonic epithelium of *villin-gp130^{Act}* mice showed more proliferation and less apoptosis during DSS colitis (Fig. 5b and Extended Data Fig. 8c), as well as elevated STAT3, SFK and YAP

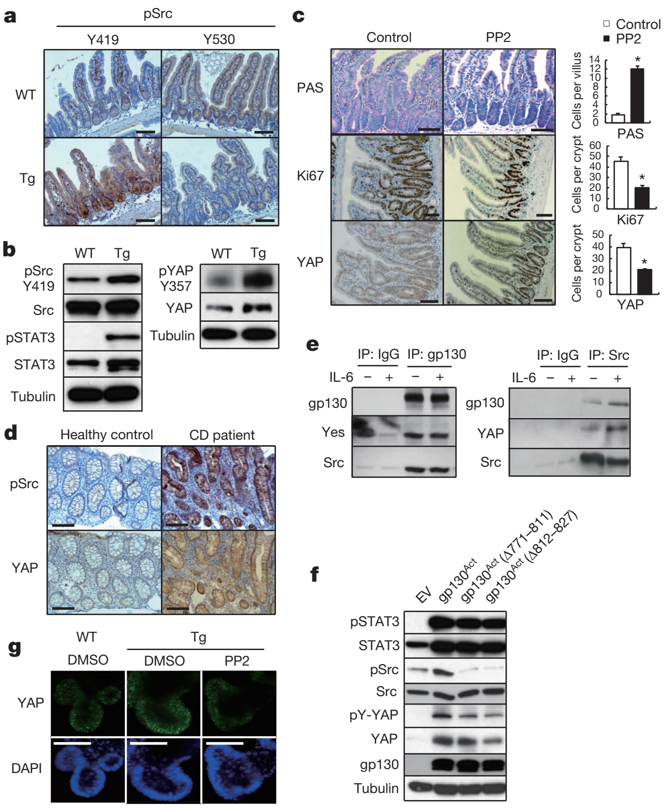


Figure 4 | SFKs activate YAP downstream to gp130^{Act} and are active in human IBD. **a**, Immunohistochemical analysis of activating (Y419) and inhibitory (Y530) Src phosphorylation in paraffin-embedded small intestinal sections from wild-type and *villin-gp130^{Act}* mice. **b**, Wild-type and *villin-gp130^{Act}* small intestinal organoids were lysed and analysed for expression and phosphorylation of the indicated proteins. **c**, *villin-gp130^{Act}* mice ($n = 4$ per group) were treated with PP2 (5 mg kg^{-1}) or vehicle once a day for 5 days. Small intestinal sections were stained as indicated. Positive cells were enumerated in each villus or crypt. Data are averages ± s.d.; $*P < 0.05$. **d**, Normal ($n = 11$) and CD ($n = 18$) human colon biopsies were fixed, sectioned and stained as indicated. Src and YAP activation were found in 11 of 18 CD specimens in areas with active disease. **e**, Co-immunoprecipitation (IP) of endogenous gp130, SFK and YAP in HT29 cells. Cells were collected with or without 2 h IL-6 (10 ng ml^{-1}) stimulation. Lysates were immunoprecipitated with gp130 (left) or Src (right) antibodies or corresponding IgG controls and probed with the indicated antibodies. **f**, Total cell lysates of T84 colon cancer cells infected with with EV, gp130^{Act}, gp130^{Act} (Δ771–811) or gp130^{Act} (Δ812–827) lentiviruses were prepared and subjected to immunoblot analysis with the indicated antibodies. **g**, Wild-type and *villin-gp130^{Act}* small intestinal organoids were treated with DMSO and PP2 (10 μM) for 24 h, stained with YAP antibody and counter stained with DAPI. Scale bars represent 100 μm (**a**, **c**, **d**, **g**) and all data are representative of at least 2–3 independent experiments.

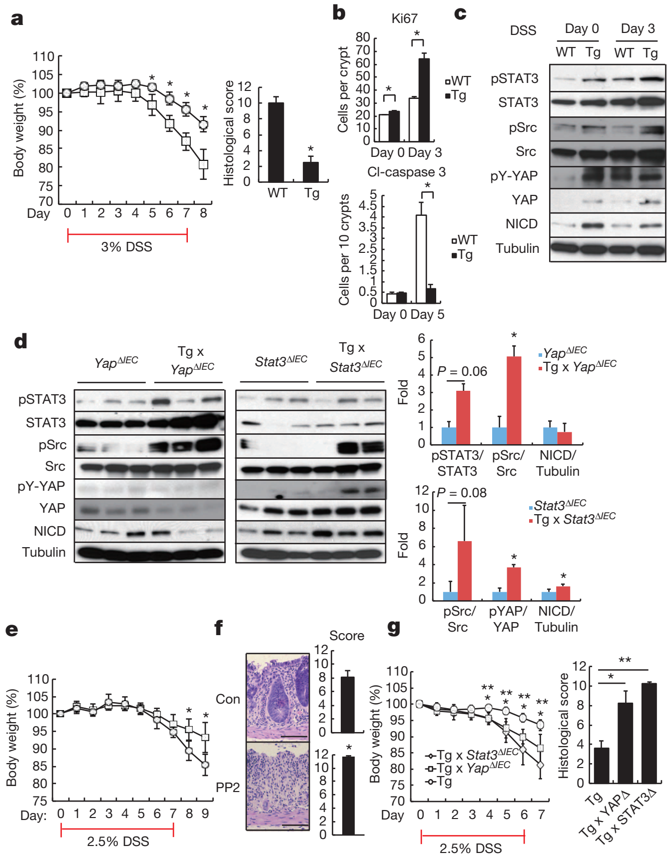


Figure 5 | gp130–SFK–YAP signalling is activated upon mucosal erosion to promote regeneration. **a**, Body weight curves and day 10 histological scores of DSS-treated wild-type (squares) and *villin-gp130^{Act}* (circles) mice ($n = 4$ per group). Results are averages ± s.d.; $*P < 0.05$. **b**, Ki67 (top panels) and cleaved-caspase 3 (bottom panels) stainings of colon sections from wild-type and *villin-gp130^{Act}* mice at day 0 and 3 (Ki67) or 5 (cleaved-caspase 3) after 3.0% DSS treatment. Positive cells were enumerated in representative microscopic fields (magnification 200× for Ki67 and 100× for cleaved-caspase 3) ($n = 6$) per time point. Results are averages ± s.e.m.; $*P < 0.05$. **c**, Wild-type and *villin-gp130^{Act}* mice were treated as above. Colonic lysates were prepared when indicated and immunoblot analysed for indicated proteins. **d**, Lysates of *Yap$^{\Delta IEC}$* and *villin-gp130^{Act}/Yap$^{\Delta IEC}$*, or *Stat3$^{\Delta IEC}$* and *villin-gp130^{Act}/Stat3$^{\Delta IEC}$* colons were prepared, and immunoblot analysed with the indicated antibodies. Data were quantified using ImageJ software and are depicted on right as averages ± s.e.m. ($n = 3$ per group); $*P < 0.05$. **e**, Body weight curves of DSS-treated control (squares) and PP2-injected (5 mg kg^{-1}) (circles) C57BL/6 mice ($n = 6$ per group). Results are averages ± s.d.; $*P < 0.05$. **f**, Mucosal histology of above mice was scored at day 10 of DSS challenge by haematoxylin and eosin staining. Results are averages ± s.e.m.; $*P < 0.05$. Scale bars represent 100 μm. **g**, Body weight curves and day 9 histological scores of *villin-gp130^{Act}* (circles), *villin-gp130^{Act}/Yap$^{\Delta IEC}$* (squares) and *villin-gp130^{Act}/Stat3$^{\Delta IEC}$* (diamonds) mice treated with 2.5% DSS ($n = 5$ per group). Results are averages ± s.d. (body weight curves) or averages ± s.e.m. (histological scores). $*P < 0.05$: *villin-gp130^{Act}* versus *villin-gp130^{Act}/Yap$^{\Delta IEC}$* mice. $**P < 0.05$: *villin-gp130^{Act}* versus *villin-gp130^{Act}/Stat3$^{\Delta IEC}$* mice.

phosphorylation and Notch activation (Fig. 5c). Notch receptors and ligands were elevated in *villin-gp130^{Act}* colons and their expression decreased upon YAP, but not STAT3, ablation (Extended Data Fig. 8d). *villin-gp130^{Act}* mice also showed an improved gut barrier function, despite no obvious differences in IEC tight junctions relative to wild-type mice (Extended Data Fig. 8e–g). DSS treatment induced IL-6 family cytokines in the colon and led to strong SFK–YAP and STAT3 activation (Extended Data Fig. 8h, i). Inhibition of SFK signalling with PP2 during DSS challenge suppressed intestinal regeneration (Fig. 5e, f and Extended Data Fig. 8j). IEC-specific YAP or STAT3 ablation in *villin-gp130^{Act}* mice increased DSS-induced weight loss and tissue damage (Fig. 5g). By contrast, gp130^{Act} conferred DSS resistance on *Yap$^{\Delta IEC}$* and *Stat3$^{\Delta IEC}$* mice, although not as strongly as its effect in wild-type mice (Extended Data Fig. 9a). Thus, both YAP and STAT3 contribute to mucosal regeneration. However, YAP was not required for STAT3 activation and STAT3 was not needed for YAP and Notch activation, and neither YAP nor STAT3 affected SFK activation (Fig. 5d). YAP, but not STAT3, was required for induction of Notch ligands and receptors, although STAT3 was essential for *Bcl2* and *Bcl-X$_L$* induction (Extended Data Fig. 9b).

We examined whether gp130 signalling controls wound healing in the absence of inflammation. Expression of *gp130^{Act}* in rat IEC6 cells enhanced wound closure after monolayer scratching (Extended Data Fig. 9c). This effect was attenuated by Src, YAP and Notch inhibitors but hardly influenced by a JAK inhibitor. Silencing of YAP also attenuated wound closure, although it did not inhibit Src activation (Extended

Data Fig. 9d, e). Together, these data support the signalling scheme outlined in Extended Data Fig. 9f and indicate that once gp130 is activated it can contribute to healing, regeneration and termination of inflammation via the SFK–YAP–Notch cascade in addition to its well-established effect on STAT3, even without ongoing inflammation.

Discussion

IL-6, produced by lamina propria myeloid cells that encounter translocating microbiota or their products, is a potent activator of the newly charted gp130–SFK–YAP–Notch pathway. Receptor binding by IL-6 and related cytokines engages gp130, which in addition to its well established effectors—the SHP2–ERK, PI(3)K–Akt–mTORC1 and JAK–STAT3 modules—interacts with and activates Src and Yes. These SFKs bind YAP through their conserved SH3 domain, which recognizes a proline-rich binding motif (SH3bm) located between the YAP WW motif and transcriptional activation domain[34], and activate YAP through Y357 phosphorylation[19]; thereby stabilizing the protein and increasing its nuclear concentration. It would be interesting to examine the effect of SH3bm inactivation on intestinal regeneration, as long as such mutations do not interfere with other aspects of YAP regulation. Another, but less direct, link between gp130 and YAP may be provided by SHP2, the subcellular distribution of which is controlled through interaction with YAP and TAZ[35].

By binding transcription factors, such as TEAD[18], YAP controls genes that stimulate cell proliferation and tissue growth and inhibit terminal differentiation[22]. Such genes include the growth factors CTGF and Areg[29] and the Notch ligand Jagged 1[36], whose expression is upregulated in the *villin-gp130^{Act}* intestinal epithelium, along with Notch 1, Notch 3 and DLL3. This results in Notch activation, which can further sustain YAP activity[37]. Although YAP and Notch are required for mucosal regeneration[16,17], the mechanisms responsible for their activation upon injury and inflammation were heretofore unknown. Until recently, activation of YAP was believed to solely depend on inhibition of Hippo signalling that retains YAP in the cytoplasm[18]. Only recently, G-protein-coupled receptors were found to activate YAP through a partially understood Hippo-independent mechanism[38], and in cancer cells Yes was demonstrated to activate YAP via tyrosine phosphorylation[19]. However, the physiological role of YAP tyrosine phosphorylation in normal tissues remained unknown. We now show that SFK-induced YAP activation, rather than Hippo inhibition, is critical for regeneration of the injured intestinal mucosa. The same pathway is activated after partial hepatectomy and may contribute to liver regeneration. YAP Y357 phosphorylation may also be regulated by the tumour suppressor RASSF1A, but its effect may be indirect as *Rassf1a$^{-/-}$* mice exhibit elevated IL-6 production after DSS-induced injury[39].

The IL-6 family member LIF stimulates self-renewal of cultured embryonic stem cells through a gp130–Yes–YAP module[40], but the *in vivo* relevance of this finding was not established. STAT3 was also suggested to maintain embryonic stem cell pluripotency[41]. Thus, it is plausible that gp130-activating cytokines are general regulators of tissue homeostasis and regeneration. Expression of IL-6 family members is elevated in a number of chronic inflammatory diseases and gastrointestinal cancers[42], and our results demonstrate frequent SFK and YAP activation in CD. Although previous efforts in targeting the pro-tumorigenic activity of IL-6-related cytokines had focused on the JAK–STAT3 module[43], it should be noted that Yes and Src are often activated in CRC, even though the cause of their activation was not identified[31,44]. Furthermore, Src controls intestinal regeneration in both mice and flies[44]. We suggest that SFK activation during IBD regeneration and CRC is caused by IL-6 family members. Nonetheless, *villin-gp130^{Act}* mice do not develop malignant tumours before 12 months of age, indicating that chronic YAP, Notch and STAT3 activation are insufficient for oncogenic transformation.

Persistent YAP and Notch activation results in skewed differentiation of intestinal stem cells, entailing expansion of immature enterocytes and under-representation of secretory cell types. The paucity of Paneth and goblet cells in *villin-gp130^{Act}* mice resembles their deficiency in IBD, which may reflect chronic elevation of IL-6 or IL-11 and persistent Src and YAP activation. Nonetheless, despite the deficiency in defensin-producing Paneth cells and mucin-producing goblet cells, *villin-gp130^{Act}* mice display an improved gut barrier function, underscoring the importance of epithelial regeneration in preventing excessive microbial translocation. Thus, future therapeutic efforts in IBD should aim at normalizing IL-6 cytokine expression, rather than complete blockade, thereby restoring immune and epithelial homeostasis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 April 2014; accepted 9 January 2015.
Published online 25 February 2015.

1. Ben-Neriah, Y. & Karin, M. Inflammation meets cancer, with NF-κB as the matchmaker. *Nature Immunol.* **12,** 715–723 (2011).
2. Medzhitov, R. Origin and physiological roles of inflammation. *Nature* **454,** 428–435 (2008).
3. Baddour, J. A., Sousounis, K. & Tsonis, P. A. Organ repair and regeneration: an overview. *Birth Defects Res. C* **96,** 1–29 (2012).
4. Johnson, R. & Halder, G. The two faces of Hippo: targeting the Hippo pathway for regenerative medicine and cancer treatment. *Nature Rev. Drug Discov.* **13,** 63–79 (2014).
5. Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140,** 883–899 (2010).
6. Neurath, M. F. New targets for mucosal healing and therapy in inflammatory bowel diseases. *Mucosal Immunol.* **7,** 6–19 (2014).
7. Garbers, C. *et al.* Plasticity and cross-talk of interleukin 6-type cytokines. *Cytokine Growth Factor Rev.* **23,** 85–97 (2012).
8. Kishimoto, T. IL-6: from its discovery to clinical applications. *Int. Immunol.* **22,** 347–352 (2010).
9. Yoshimura, A., Naka, T. & Kubo, M. SOCS proteins, cytokine signalling and immune regulation. *Nature Rev. Immunol.* **7,** 454–465 (2007).
10. Putoczki, T. & Ernst, M. More than a sidekick: the IL-6 family cytokine IL-11 links inflammation to cancer. *J. Leukoc. Biol.* **88,** 1109–1117 (2010).
11. Rose-John, S., Mitsuyama, K., Matsumoto, S., Thaiss, W. M. & Scheller, J. Interleukin-6 trans-signaling and colonic cancer associated with inflammatory bowel disease. *Curr. Pharm. Des.* **15,** 2095–2103 (2009).
12. Pilati, C. *et al.* Somatic mutations activating STAT3 in human inflammatory hepatocellular adenomas. *J. Exp. Med.* **208,** 1359–1366 (2011).
13. Grivennikov, S. *et al.* IL-6 and Stat3 are required for survival of intestinal epithelial cells and development of colitis-associated cancer. *Cancer Cell* **15,** 103–113 (2009).
14. Rebouissou, S. *et al.* Frequent in-frame somatic deletions activate gp130 in inflammatory hepatocellular tumours. *Nature* **457,** 200–204 (2009).
15. Tanaka, T., Narazaki, M. & Kishimoto, T. Therapeutic targeting of the interleukin-6 receptor. *Annu. Rev. Pharmacol. Toxicol.* **52,** 199–219 (2012).
16. Cai, J. *et al.* The Hippo signaling pathway restricts the oncogenic potential of an intestinal regeneration program. *Genes Dev.* **24,** 2383–2388 (2010).
17. Okamoto, R. *et al.* Requirement of Notch activation during regeneration of the intestinal epithelia. *Am. J. Physiol. Gastrointest. Liver Physiol.* **296,** G23–G35 (2009).
18. Yu, F. X. & Guan, K. L. The Hippo pathway: regulators and regulations. *Genes Dev.* **27,** 355–371 (2013).
19. Rosenbluh, J. *et al.* β-Catenin-driven cancers require a YAP1 transcriptional complex for survival and tumorigenesis. *Cell* **151,** 1457–1473 (2012).
20. Bray, S. J. Notch signalling: a simple pathway becomes complex. *Nature Rev. Mol. Cell Biol.* **7,** 678–689 (2006).
21. Madison, B. B. *et al.* Cis elements of the villin gene control expression in restricted domains of the vertical (crypt) and horizontal (duodenum, cecum) axes of the intestine. *J. Biol. Chem.* **277,** 33275–33283 (2002).
22. Camargo, F. D. *et al.* YAP1 increases organ size and expands undifferentiated progenitor cells. *Curr. Biol.* **17,** 2054–2060 (2007).
23. Zhou, D. *et al.* Mst1 and Mst2 protein kinases restrain intestinal stem cell proliferation and colonic tumorigenesis by inhibition of Yes-associated protein (Yap) overabundance. *Proc. Natl Acad. Sci. USA* **108,** E1312–E1320 (2011).
24. Fre, S. *et al.* Notch signals control the fate of immature progenitor cells in the intestine. *Nature* **435,** 964–968 (2005).
25. van Es, J. H. *et al.* Notch/γ-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* **435,** 959–963 (2005).
26. Sato, T. & Clevers, H. Growing self-organizing mini-guts from a single intestinal stem cell: mechanism and applications. *Science* **340,** 1190–1194 (2013).
27. Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469,** 415–418 (2011).
28. Murakami, D. *et al.* Presenilin-dependent gamma-secretase activity mediates the intramembranous cleavage of CD44. *Oncogene* **22,** 1511–1516 (2003).
29. Zhang, J. *et al.* YAP-dependent induction of amphiregulin identifies a non-cell-autonomous component of the Hippo pathway. *Nature Cell Biol.* **11,** 1444–1450 10.1038/ncb1993 (2009).

30. Nishioka, N. *et al.* The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Dev. Cell* **16,** 398–410 (2009).
31. Chen, J., Elfiky, A., Han, M., Chen, C. & Saif, M. W. The role of Src in colon cancer and its therapeutic implications. *Clin. Colorectal Cancer* **13,** 5–13 (2014).
32. Levy, D., Adamovich, Y., Reuven, N. & Shaul, Y. Yap1 phosphorylation by c-Abl is a critical step in selective activation of proapoptotic genes in response to DNA damage. *Mol. Cell* **29,** 350–361 (2008).
33. Azzolin, L. *et al.* YAP/TAZ incorporation in the β-catenin destruction complex orchestrates the Wnt response. *Cell* **158,** 157–170 (2014).
34. Sudol, M. Yes-associated protein (YAP65) is a proline-rich phosphoprotein that binds to the SH3 domain of the Yes proto-oncogene product. *Oncogene* **9,** 2145–2152 (1994).
35. Tsutsumi, R. *et al.* YAP and TAZ, Hippo signaling targets, act as a rheostat for nuclear SHP2 function. *Dev. Cell* **26,** 658–665 (2013).
36. Tschaharganeh, D. F. *et al.* Yes-associated protein up-regulates Jagged-1 and activates the Notch pathway in human hepatocellular carcinoma. *Gastroenterology* **144,** 1530–1542 (2013).
37. Li, Y., Hibbs, M. A., Gard, A. L., Shylo, N. A. & Yun, K. Genome-wide analysis of N1ICD/RBPJ targets *in vivo* reveals direct transcriptional regulation of Wnt, SHH, and hippo pathway effectors by Notch1. *Stem Cells* **30,** 741–752 (2012).
38. Yu, F. X. *et al.* Regulation of the Hippo-YAP pathway by G-protein-coupled receptor signaling. *Cell* **150,** 780–791 (2012).
39. Gordon, M. *et al.* The tumor suppressor gene, RASSF1A, is essential for protection against inflammation-induced injury. *PLoS ONE* **8,** e75483 (2013).
40. Tamm, C., Bower, N. & Anneren, C. Regulation of mouse embryonic stem cell self-renewal by a Yes-YAP-TEAD2 signaling pathway downstream of LIF. *J. Cell Sci.* **124,** 1136–1144 (2011).
41. Raz, R., Lee, C. K., Cannizzaro, L. A., d'Eustachio, P. & Levy, D. E. Essential role of STAT3 for embryonic stem cell pluripotency. *Proc. Natl Acad. Sci. USA* **96,** 2846–2851 (1999).
42. Taniguchi, K. & Karin, M. IL-6 and related cytokines as the critical lynchpins between inflammation and cancer. *Semin. Immunol.* **26,** 54–74 (2014).
43. Hedvat, M. *et al.* The JAK2 inhibitor AZD1480 potently blocks Stat3 signaling and oncogenesis in solid tumors. *Cancer Cell* **16,** 487–497 (2009).
44. Cordero, J. B. *et al.* c-Src drives intestinal regeneration and transformation. *EMBO J.* **33,** 1474–1491 (2014).

Acknowledgements We thank D. Pan and S. Akira for *Yap*$^{fl/fl}$ and *Stat3*$^{fl/fl}$ mice, respectively. We also thank D. L. Gumucio for a plasmid containing the 12.4-kb villin promoter, T. Sato, H. Clevers and Y. Hippo for protocols describing intestinal organoid culture, C. Kuo for R-spondin1-producing cells, D. Huszar for AZD1480, F. Schaper for plasmids, L. Eckmann for advice, A. Umemura, H. Nakagawa, H. Ogata, E. J. Park, G. Y. Yu, J. Font-Burgada, D. Dhar, J. Kim and E. Seki for providing liver samples, J. Zhao, T. Meerloo, Y. Jones, L. Gapuz, R. Ly, N. Varki, D. Aki, N. Hiramatsu, T. Moroishi, Y. Endo, H. Nishinakamura, A. Chang and T. Lee for technical advice and assistance, and Cell Signaling, Santa Cruz Biotechnology and GeneTex for antibodies. This work was supported by Postdoctoral Fellowship for Research Abroad and Research Fellowship for Young Scientists from the Japan Society for the Promotion of Science, a Uehara Memorial Foundation Fellowship, the Mochida Memorial Foundation for Medical and Pharmaceutical Research, and the Kanae Foundation for the Promotion of Medical Science to K.T.; a traveling grant NSC-101-2918-I-006-005 and a research grant NSC-103-2320-B-006-032 by National Science Council of Taiwan to L.-W.W.; NIH R00DK088589, FCCC-Temple University Nodal grant, AACR-Landon Innovator Award in Tumor Microenvironment, and the Pew Scholar in Biomedical Sciences Program for S.I.G.; a CCFA fellowship (RFA2927) to P.R.d.J.; Croucher Foundation and China Postdoctoral Science Foundation to K.W.; by the Research Service of the Department of Veterans Affairs to S.B.H.; by the NIH and the UCSD Digestive Disease Research Center Grant to J.T.C. and W.J.S.; by the NIH EY022611 and CA132809 to K.-L.G.; and by the NIH CA118165-09 and AACR to M.K., who is an American Cancer Society Research Professor and holds the Ben and Wanda Hildyard Chair for Mitochondrial and Metabolic Diseases.

Author Contributions K.T. and M.K. conceived the project. K.T., L.-W.W., S.I.G., P.R.d.J., I.L., F.-X.Y., K.W., G.H. performed the experiments. K.T., L.-W.W., P.R.d.J, G.H. and M.K. analysed data. J.Z.-R. provided gp130 mutants, S.B.H., J.T.C., B.S.B. and W.J.S. provided human specimens, S.I.G., E.R., Y.M., A.Y., J.Z.-R. and K.-L.G. provided conceptual advice. K.T., L.-W.W. and M.K. wrote the manuscript, with all authors contributing to the writing and providing advice.

Author Information Microarray data reported here have been deposited in the ArrayExpress database under accession E-MEXP-E-MTAB-2400. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.K. (karinoffice@ucsd.edu).

METHODS

Mice. gp130^{Act} cDNA[14] was amplified by PCR and subcloned into a plasmid containing the 12.4-kb villin promoter (A gift from D. L. Gumucio, University of Michigan)[21]. The 15.7-kb expression cassette was excised by PmeI digestion, purified, and injected into fertilized C57BL/6 oocytes to obtain founder mice, two of which transmitted the *gp130^{Act}* transgene. C57BL/6 mice were purchased from Charles River Laboratories. *villin-cre* mice were obtained from The Jackson Laboratory. *Stat3$^{fl/fl}$* and *Yap$^{fl/fl}$* mice have been described[45,46]. Two- to three-month-old sex-matched mice were used for all experimental procedures. All mice were on the C57BL/6 background and were maintained in filter-topped cages on autoclaved food and water at University of California, San Diego (UCSD) according to NIH guidelines. Bedding was interchanged between the different strains to minimize microbiome alterations. All experimental procedures were reviewed and approved by the UCSD Institutional Animal Care and Use Committee and all experiments were performed in accordance with UCSD and NIH guidelines and regulations.

Human colon samples. Human tissue specimens were retrospectively obtained from routine colonoscopic biopsies fixed in buffered formalin and embedded in paraffin using standard methods. Corresponding clinical data was recorded from medical records and de-identified. The study was approved by the institutional review board and research and development committee of the VA San Diego Healthcare System. Subjects had either established Crohn's disease based on clinical, endoscopic, and histopathological criteria, or were normal based on the absence of systemic clinical disease, a normal colonoscopic exam, and normal histopathology.

Reagents and plasmids. Recombinant Noggin, IL-6 and IL-11 were purchased from Peprotech, recombinant EGF and wheat germ agglutinin conjugated with Alexa Fluor 488 were from Life Technologies, DBZ (γ-secretase inhibitor) from Axon Medchem, PP2 (Src inhibitor), SU 6656 (Src inhibitor) and LY294002 (PI3K inhibitor) from Sigma, PD0325901 (MEK inhibitor) from Stemgent, and AZD0530 (Src inhibitor) and verteporfin (YAP inhibitor) from Selleck Chemicals. AZD1480 (JAK inhibitor) was provided by D. Huszar (AstraZeneca). 293T-HA-RspoI-Fc cells were provided by C. Kuo (Stanford University)[47]. To generate gp130 expression vectors, wild-type gp130, gp130^{Act} and gp130^{SA} were amplified by PCR and subcloned into lentiviral expression vectors. gp130^{Act} deletion mutants were constructed by a PCR-based approach using PrimeSTAR Max (TaKaRa Bio). pcDNA Flag–YAP[32] was obtained from Addgene (Plasmid 18881). For shRNA transduction, pLKO.1-puro lentiviral vectors (Sigma-Aldrich) targeting YAP (TRCN0000238432) or luciferase (control) were used.

Cell culture. HEK293T, HT29, HCA7, SW480, DLD1, HCT116 and IEC6 cells were cultured in Dulbecco's modified Eagle' medium (DMEM) supplemented with 10% fetal bovine serum (FBS), penicillin and streptomycin. T84 cells were cultured in DMEM/F12 supplemented with 10% FBS, penicillin and streptomycin.

Cytokine treatment. Following overnight starvation with serum-free medium, colon cancer cells were treated with recombinant human IL-6 or IL-11 for the indicated time. For those treated with inhibitors, PP2 or AZD1480, cells were pretreated for 1–2 h with the inhibitor before the cytokine treatment.

Transient transfection. HEK293T cells were transfected with CA-Src expression vector[48], or co-transfected with the indicated plasmids, gp130 expression vector together with c-Src expression vector[48] by using Lipofectamine 2000 (Invitrogen) as described by the manufacturer. Forty-eight hours later, total protein lysates were harvested from the transfected cells and subjected to immunoprecipitation and immunoblot analysis.

Ectopic expression of gp130 variants in colon cancer cells. Cultured cells were infected with empty viruses (EV), or lentiviruses bearing Flag-tagged wild-type (WT), active (gp130^{Act}) or superactive gp130 (gp130^{SA}). Stably infected cells were enriched by G418 selection for 2–3 weeks. Ectopic gp130 expression was confirmed by immunoblot analysis using anti-Flag and anti-gp130 antisera.

Nuclear and cytosol fractionation. Nuclear and cytosol fractions were prepared as previously described[49]. Briefly, the indicated cells were washed with PBS and then harvested in lysis buffer (10 mM Tris-HCl, pH 6.8, 10 mM NaCl, 3 mM $MgCl_2$, 0.05% NP40, 1 mM EGTA, 1 mM Na_3VO_4, 50 mM NaF and 1 nM okadaic acid) containing protease inhibitors. Following lysate centrifugation at 20,800*g* for 15 min at 4 °C, the supernatant was collected for cytosolic fraction. The pellet was washed with buffers I and II (I: 10 mM HEPES, pH 6.8, 25 mM NaCl, 3 mM $MgCl_2$, 300 mM sucrose, 1 mM EGTA, 1 mM Na_3VO_4 and 50 mM NaF; II: 1 M sucrose, 1 mM $NaVO_4$ and 50 mM NaF) followed by centrifugation at 2,700*g* for 5 min at 4 °C. The resulting pellet was extracted with an extraction buffer (20 mM HEPES, pH 7.9, 300 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1 mM Na_3VO_4, 0.1 mM β-glycerophosphate, 50 mM NaF and 1 nM okadaic acid) on ice for 30 min. After centrifugation at 20,800*g* for 15 min at 4 °C, the supernatant was saved as nuclear extract for subsequent studies.

Immunoprecipitation and immunoblot analysis. Equal amounts of total protein from each sample were fractionated by SDS–PAGE and blotted onto polyvinylidene difluoride membrane. Protein blots were hybridized with the indicated primary antibody of interest and then with secondary antibody, followed by detection with Immobilon Western system (Millipore Corp.). For immunoprecipitation, the cells were lysed in a buffer containing 50 mM Tris-HCl (pH 7.6), 150 mM NaCl, 0.1% SDS, 0.1% sodium deoxycholate, 1% Triton X-100, PhosSTOP Phosphatase Inhibitor Cocktail (Roche Applied Sciences), and Complete Protease Inhibitor Cocktail (Roche Applied Sciences). The lysates were cleared by centrifugation at top speed for 15 min. Immune complexes were collected with the antibody of interest and protein G/A agarose beads, followed by immunoblotting as described. Immunoblotting was performed with antibodies to phospho-YAP Y357 (Abcam), c-Yes, β-catenin (BD), phospho-Src, Src, phospho-STAT3, phospho-ERK1/2, phospho-Akt S473, phospho-YAP S127, YAP, TAZ, NICD, active-β-catenin, phospho-LATS, LATS2 (Cell Signaling), phospho-STAT1 (Upstate), HES1, STAT3, STAT1, ERK2, Akt, lamin A, HDAC1, GAPDH (Santa Cruz), tubulin, actin, Flag (Sigma) and CTGF (GeneTex).

Real-Time PCR analysis. Total RNA was extracted using TRIzol (Invitrogen) or an RNeasy Plus kit (Qiagen). RNA was reverse transcribed using an iScript cDNA Synthesis kit (Bio-Rad) or SuperScript VILO cDNA Synthesis kit (Life Technologies). Real-time PCR was performed using SYBR green (Bio-Rad) on a Bio-Rad CFX96 machine. The comparative threshold cycle method and an internal control (18S rRNA, β-Actin or GAPDH) were used to normalize the expression of the target genes.

***In vivo* treatment with inhibitors.** Mice were treated with DBZ (10 μmol kg^{-1}), PP2 (5 mg kg^{-1}), PD0325901 (25 mg kg^{-1}), LY294002 (100 mg kg^{-1}) and vehicle for 5 days except DSS-induced colitis as previously described[25,50–52]. The mice were killed after the treatment and the intestines were collected for further analysis. For DSS-induced colitis, mice were treated with PP2 (5 mg kg^{-1}) or vehicle every day from day −2 to day 10. Mice were treated with PP2 (5 mg kg^{-1}) or vehicle every day from day −1 to day 2 in a model of partial hepatectomy. All experiments were performed in a randomized manner by investigators blinded to the groups.

***In vitro* scratch/wound healing assay.** IEC6 cells infected with EV or gp130^{Act} lentiviruses were plated to confluence and starved overnight, and then scratched to cause a wound in the monolayer. Cells were treated with DMSO, PP2 (10 μM), AZD1480 (1 μM), verteporfin (1 μg ml^{-1}) or DBZ (10 μM) for 24 h. Images from five independent fields were taken at 0- and 24-h time points with a Carl Zeiss inverted microscope at a magnification of 10×. The percentage of wound closure was quantified with Image J software.

DSS-induced colitis. Mice received water with 2.0–3.0% dextran sulfate sodium (DSS; MP Biomedicals, molecular weight 36,000–50,000 kDa, or Affymetrix, molecular weight 40,000–50,000 kDa) for 6–7 days. After this, mice were maintained on regular water and were killed on days 8–10. One half of the distal colon was taken as a tissue sample and snap-frozen for subsequent RNA and protein analysis. The other half was fixed with 4% paraformaldehyde or 10% neutral buffered formalin for 24 h for subsequent paraffin embedding and histological analysis. The clinical course of disease was followed daily by measurement of body weight. Colonic histology scores were determined by an observer blinded to genotype as previously described[53].

Histological analysis. Mouse intestine samples were fixed with 4% paraformaldehyde or 10% neutral buffered formalin for 24 h and paraffin embedded. 5-μm-thick serial sections were used for H&E or other staining. Antibodies used were against: CD44 C-terminal, Chromogranin A (Abcam), BrdU, CD45 (eBioscience), phospho-Src, phospho-STAT3, phospho-STAT1, YAP, HES1, MMP7, phospho-ERK1/2, phospho-S6, cyclin D1 (Cell Signaling), lysozyme (Santa Cruz), Ki67 and c-Myc (GeneTex). Measurements for each quantitative outcome were collected from 30–50 crypts or villi analysed from 3–6 independent fields of small intestine or colon of several independent mice ($n = 2$–6).

Electron microscopic analysis. Tissue samples from small intestines were fixed in modified Karnovsky's fixative (2.5% glutaraldehyde and 2% paraformaldehyde in 0.15 M sodium cacodylate buffer, pH 7.4) for 24 h at 4 °C, post-fixed in 1% osmium tetroxide in 0.15 M cacodylate buffer for 1 h and stained en bloc in 2% uranyl acetate for 1 h. Samples were dehydrated in ethanol, embedded in Durcupan epoxy resin (Sigma-Aldrich), sectioned at 50 to 60 nm on a Leica UCT ultramicrotome, and picked up on Formvar and carbon-coated copper grids. Sections were stained with 2% uranyl acetate for 5 min and Sato's lead stain for 1 min. Grids were viewed using a JEOL 1200EX II (JEOL, Peabody, MA) transmission electron microscope and photographed using a Gatan digital camera (Gatan, Pleasanton, CA).

Intestinal organoid culture and staining. Small intestinal crypts were isolated from wild-type and *villin-gp130^{Act}* small intestines and were cultured and stained as previously described[54,55].

Intestinal permeability *in vivo*. Intestinal permeability and gut barrier function were measured using the FITC-labelled dextran method as previously described[56] and by measuring fecal albumin. Fecal albumin measurements were performed with dried fecal pellets using the Mouse Albumin ELISA Quantitation Set obtained from Bethyl Laboratories (Montgomery, TX, USA) according to manufacturer's instructions.

***In situ* hybridization.** Intestinal paraffin sections were hybridized to anti-Cryptdin-1 probe as previously described[57].

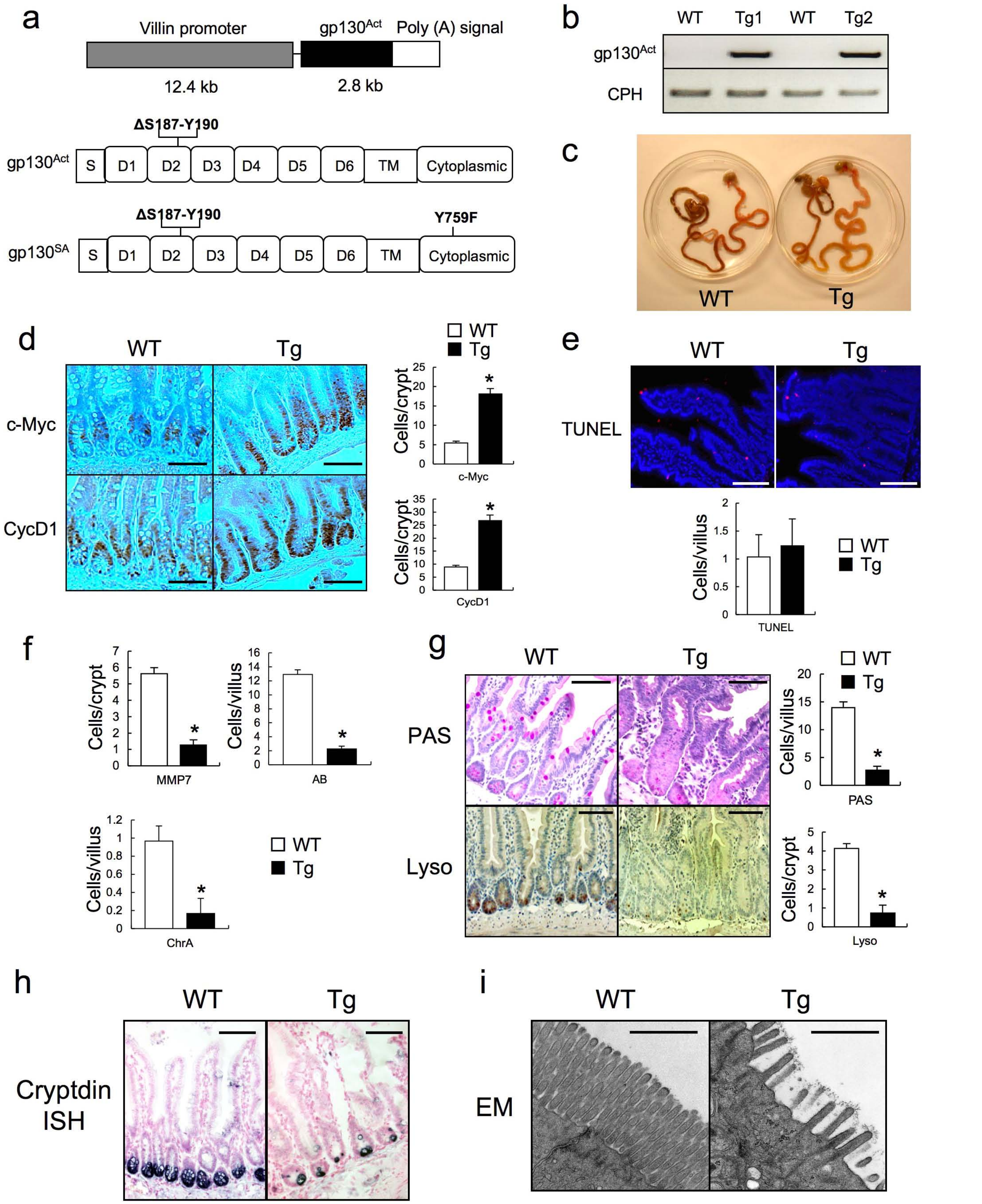
Isolation of primary mouse hepatocytes and two-thirds partial hepatectomy. Primary mouse hepatocytes were isolated and cultured as described[58]. Two-thirds partial hepatectomy was performed as previously described[59].

Microarray analysis. Small intestinal crypts were isolated from wild-type and *villin-gp130^{Act}* small intestines as previously described[54]. Total RNA was isolated from the isolated crypts using the RNeasy Mini Kit (Qiagen) and used for microarray analysis with the Illumina MouseWG-6 v2 Expression BeadChip (Illumina, San Diego). Microarray processing, data normalization and analysis were done as previously described[60].

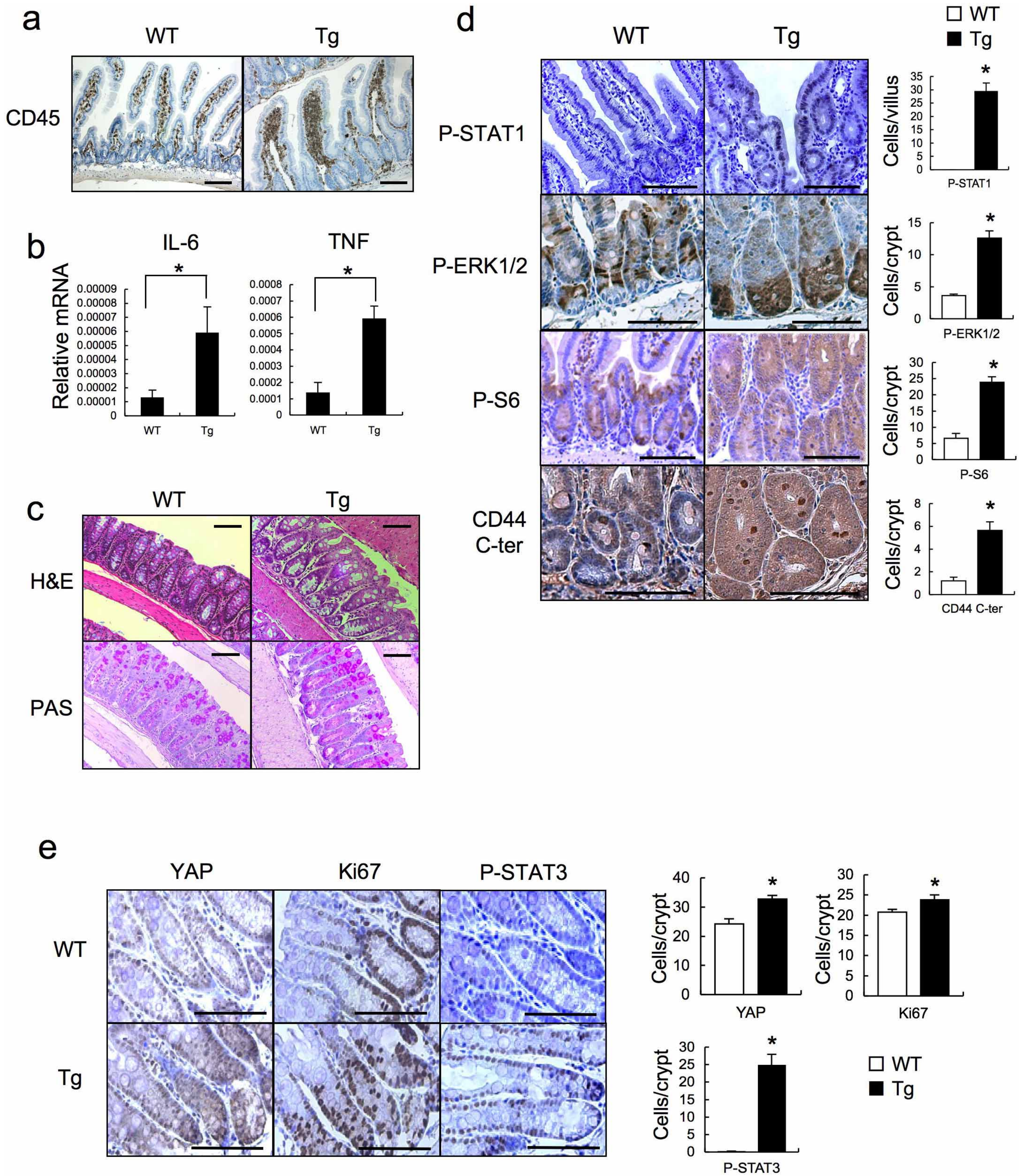
Accession codes. Microarray data reported here have been deposited in the ArrayExpress database (accession E-MEXP-E-MTAB-2400).

Statistical analysis. Data are expressed as the mean ± s.d. or the mean ± s.e.m. Statistical analysis was conducted using Student's *t*-test or one-way ANOVA followed by the Tukey–Kramer test for multiple comparisons. Fisher's exact test was used for comparison of categorical variables between CD patients and healthy controls. No statistical methods were used to predetermine sample size. Statistical significance was defined as a P value of <0.05.

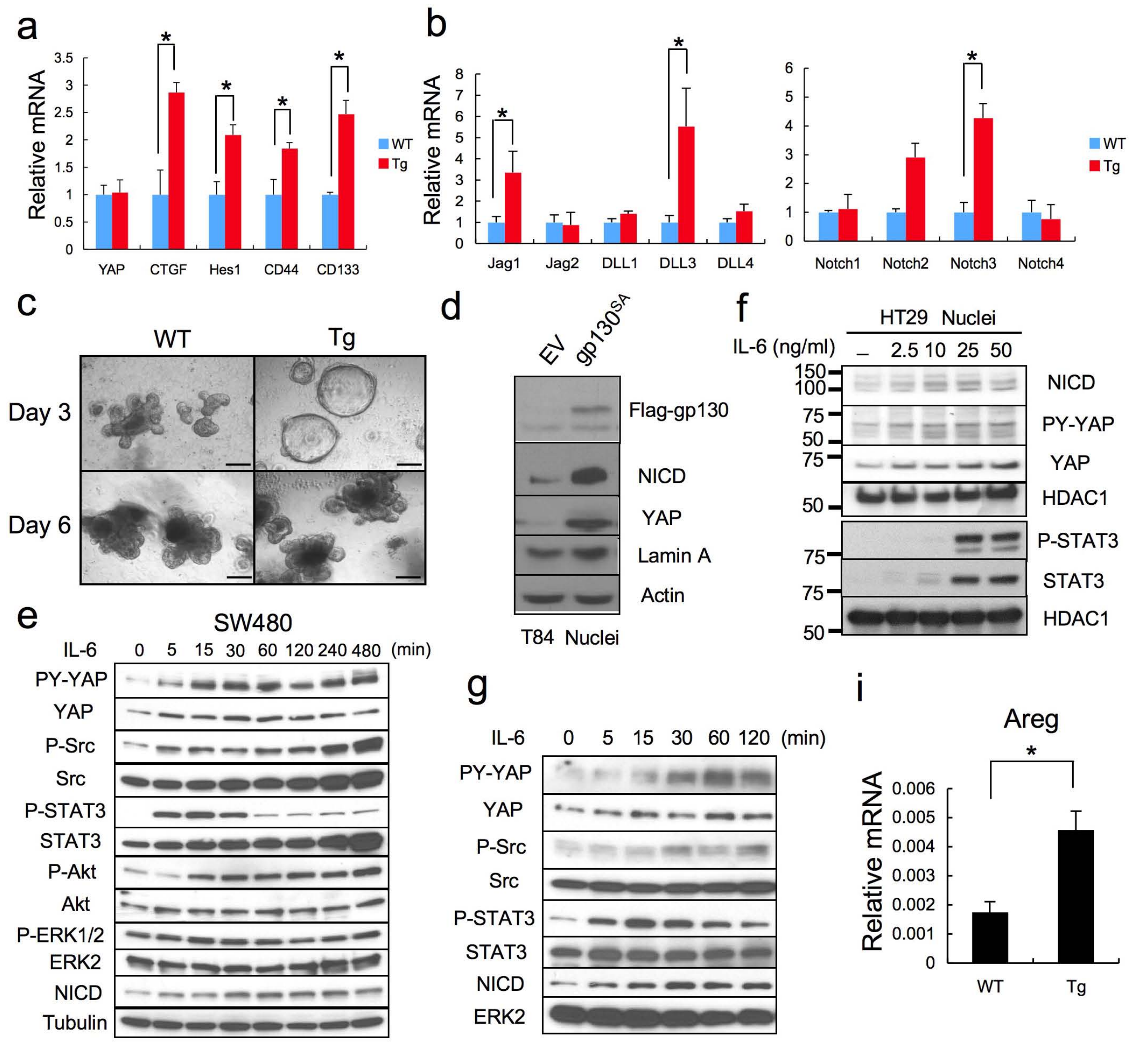
45. Takeda, K. *et al.* Enhanced Th1 activity and development of chronic enterocolitis in mice devoid of Stat3 in macrophages and neutrophils. *Immunity* **10,** 39–49 (1999).
46. Zhang, N. *et al.* The Merlin/NF2 tumor suppressor functions through the YAP oncoprotein to regulate tissue homeostasis in mammals. *Dev. Cell* **19,** 27–38 (2010).
47. Ootani, A. *et al.* Sustained *in vitro* intestinal epithelial culture within a Wnt-dependent stem cell niche. *Nature Med.* **15,** 701–706 (2009).
48. Holzer, R. G. *et al.* Saturated fatty acids induce c-Src clustering within membrane subdomains, leading to JNK activation. *Cell* **147,** 173–184 (2011).
49. Hu, H. T. *et al.* Tie2-R849W mutant in venous malformations chronically activates a functional STAT1 to modulate gene expression. *J. Invest. Dermatol.* **128,** 2325–2333 10.1038/jid.2008.89 (2008).
50. Nam, J. S., Ino, Y., Sakamoto, M. & Hirohashi, S. Src family kinase inhibitor PP2 restores the E-cadherin/catenin cell adhesion system in human cancer cells and reduces cancer metastasis. *Clin. Cancer Res.* **8,** 2430–2436 (2002).
51. Lee, S. H. *et al.* ERK activation drives intestinal tumorigenesis in *Apc$^{min/+}$* mice. *Nature Med.* **16,** 665–670 (2010).
52. Hu, L., Zaloudek, C., Mills, G. B., Gray, J. & Jaffe, R. B. *In vivo* and *in vitro* ovarian carcinoma growth inhibition by a phosphatidylinositol 3-kinase inhibitor (LY294002). *Clin. Cancer Res.* **6,** 880–886 (2000).
53. Katakura, K. *et al.* Toll-like receptor 9-induced type I IFN protects mice from experimental colitis. *J. Clin. Invest.* **115,** 695–702 (2005).
54. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459,** 262–265 (2009).
55. Sato, T. *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141,** 1762–1772 (2011).
56. Guma, M. *et al.* Constitutive intestinal NF-κB does not trigger destructive inflammation unless accompanied by MAPK activation. *J. Exp. Med.* **208,** 1889–1900 (2011).
57. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449,** 1003–1007 (2007).
58. Umemura, A. *et al.* Liver damage, inflammation, and enhanced tumorigenesis after persistent mTORC1 inhibition. *Cell Metab.* **20,** 133–144 (2014).
59. Mitchell, C. & Willenbring, H. A reproducible and well-tolerated method for 2/3 partial hepatectomy in mice. *Nature Protocols* **3,** 1167–1170 (2008).
60. Kozak, I. *et al.* A degenerative retinal process in HIV-associated non-infectious retinopathy. *PLoS ONE* **8,** e74712 (2013).

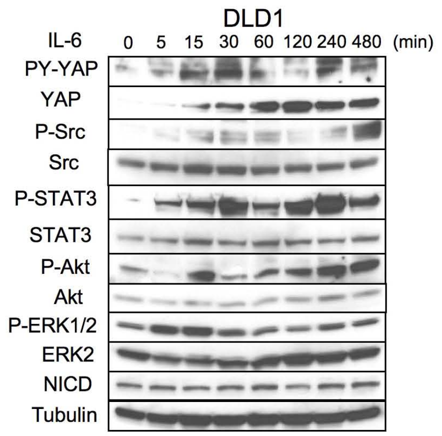


Extended Data Figure 1 | gp130^{Act} expression and intestinal phenotype. **a**, Schematic diagram of the *villin-gp130^{Act}* transgenic (Tg) construct and the gp130^{Act} and gp130^{SA} variants. **b**, Expression of *gp130^{Act}* in the *villin-gp130^{Act}* jejunum was confirmed by RT–PCR with specific primers for human *gp130*. Cyclophilin (*CPH*) was used as an internal control. **c**, Representative images of wild-type (WT) and *villin-gp130^{Act}* intestines. c-Myc and cyclin D1 (**d**) and TUNEL (red, TUNEL; blue, DAPI) (**e**) staining of paraffin-embedded sections of wild-type and *villin-gp130^{Act}* small intestines from 3-month-old mice. Positive cells were enumerated in each villus or crypt ($n = 6$). Data represent averages ± s.d.; $*P < 0.05$. **f**, MMP7-, AB- and ChrA-positive cells in wild-type and *villin-gp130^{Act}* small intestines were enumerated in each villus or crypt ($n = 6$). Data represent averages ± s.d.; $*P < 0.05$. **g**, Paraffin-embedded sections of wild-type and *villin-gp130^{Act}* small intestines were analysed by PAS and lysozyme staining. Positive cells were enumerated in each villus or crypt ($n = 6$). Data represent averages ± s.d.; $*P < 0.05$. **h**, Cryptdin mRNA in wild-type and *villin-gp130^{Act}* jejunum was detected by *in situ* hybridization. **i**, Transmission electron microscopy (TEM) of the apical surface of wild-type and transgenic small intestines. Scale bars represent 100 μm (**d**, **e**, **g**, **h**) and 1 μm (**i**) and all data are representative of at least 2–3 independent experiments.



Extended Data Figure 2 | Aberrant intestinal differentiation and activation of gp130 effectors in *gp130*^{Act} mice. **a**, Paraffin-embedded sections of wild-type and *villin-gp130*^{Act} (Tg) small intestines were analysed by CD45 staining. **b**, Lysates of wild-type and *villin-gp130*^{Act} jejuna were prepared, and expression of *Il6* and *Tnf* mRNAs was analysed by qRT–PCR ($n = 3$). Results are averages ± s.e.m.; $*P < 0.05$. **c**, H&E and PAS staining of paraffin-embedded sections of wild-type and *villin-gp130*^{Act} large intestines. **d**, pSTAT1, pERK1/2, pS6 and CD44 C-terminal stainings of paraffin-embedded sections of wild-type and *villin-gp130*^{Act} small intestines. Positive cells were enumerated in each villus or crypt ($n = 4$). Data represent averages ± s.d.; $*P < 0.05$. **e**, YAP, Ki67 and pSTAT3 stainings of paraffin-embedded sections of wild-type and *villin-gp130*^{Act} large intestines. Positive cells were enumerated in each crypt ($n = 4$). Data represent averages ± s.d.; $*P < 0.05$. Scale bars represent 100 μm (**a, c-e**) and all data are representative of at least 2–3 independent experiments.

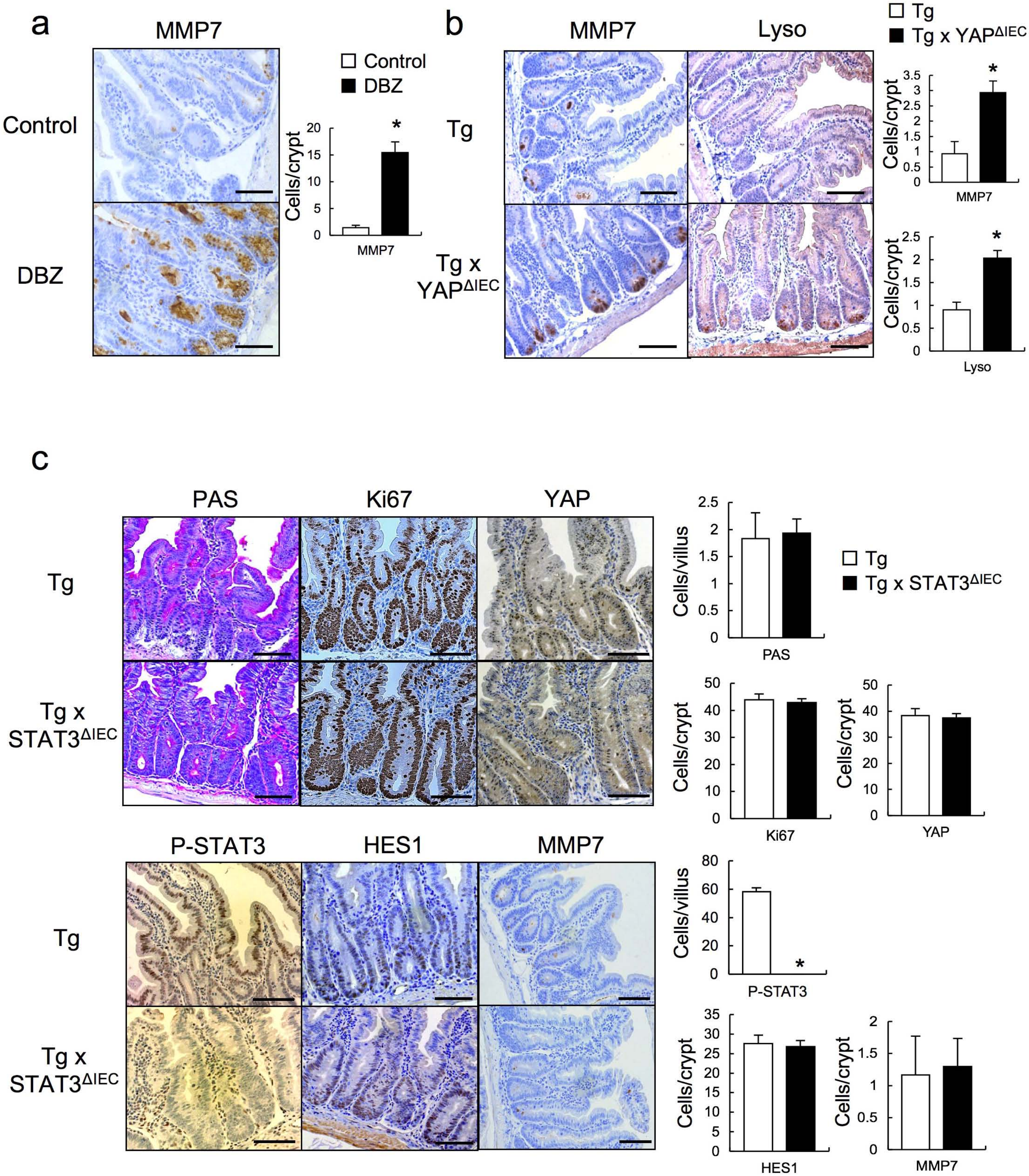




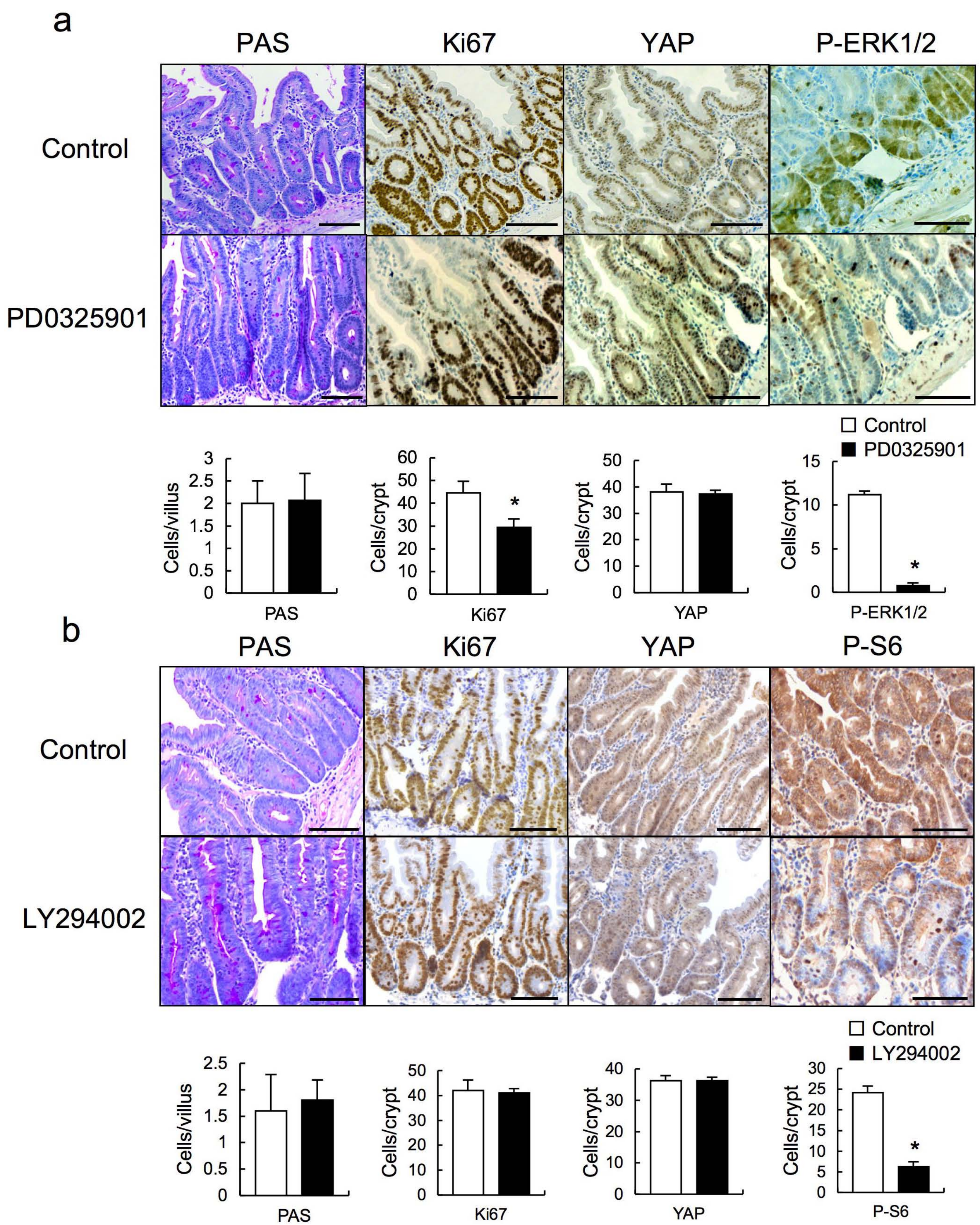
h

Gene	Fold induction	Gene	Fold induction	Gene	Fold induction
Ifi27	73.46	SOCS1	1.27	Sox9	-1.48
Ifit2	43.97	STAT1	3.23	Cyp2b10	-40.09
Oas2	36.94	SOCS3	2.05	CD74	-10.05
Ifit3	27.25	STAT3	1.31	H2-Ab1	-14.96
Mx2	17.21	SAA1	1.49	Ly6a	66.20
TRAFD1	6.37	Jag1	1.35	TGFbR2	2.69
IRF7	5.74	ADAM17	1.38	CD44	1.48
MyD88	1.72	Notch4	1.64	HIF1a	1.64
Cxcl10	3.08	BMI1	1.28	Defa1	-2.22
Ccl5	-15.53	Lgr5	-1.72	Areg	11.16

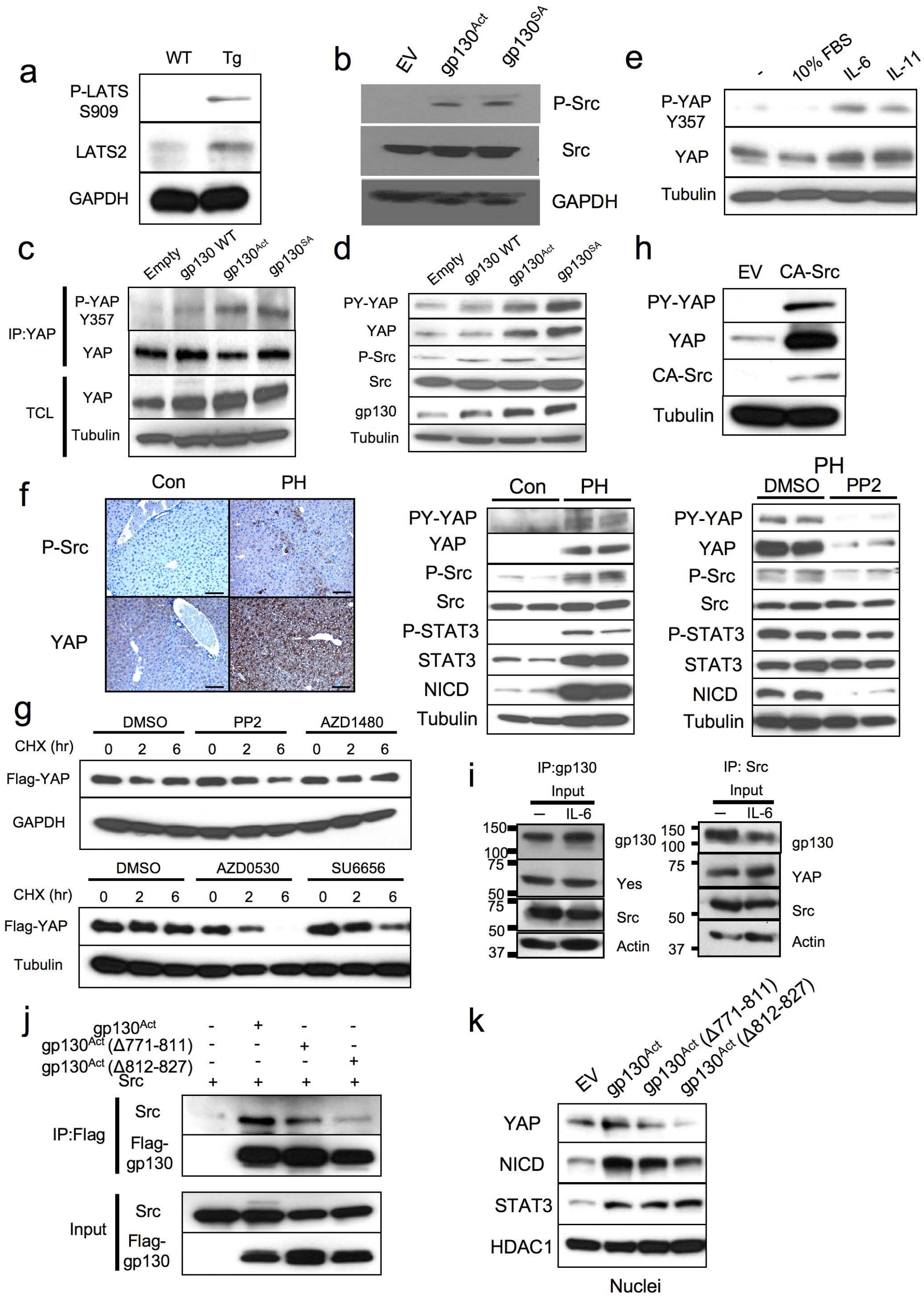
Extended Data Figure 3 | IL-6 and gp130 induce Notch and YAP activation in intestinal organoids and cancer cells, and gene expression analysis of intestinal crypts. **a**, **b**, Wild-type and *villin-gp130*^{Act} organoids were cultured, their RNA extracted, and expression of the indicated mRNA species was measured by qRT–PCR ($n = 3$). Results are averages ± s.e.m.; $*P < 0.05$. **c**, Appearance of wild-type and *villin-gp130*^{Act} small intestinal organoids cultured in standard EGF/Noggin/R-spondin 1 medium. **d**, Nuclei of T84 colon cancer cells transfected with either empty vector (EV) or a vector encoding superactive gp130 (gp130^{SA}) were lysed and subjected to immunoblot analysis with the indicated antibodies. Lamin A, a nuclear marker. Actin, a loading control. **e**, Lysates of serum-starved SW480 (upper) or DLD1 (lower) colon cancer cells treated for 0–480 min with IL-6 at 50 ng ml^{-1} were subjected to immunoblot analysis using the indicated antibodies. **f**, Nuclei of serum-starved HT29 colon cancer cells treated for 24 h with IL-6 at 0–50 ng ml^{-1} were lysed and subjected to immunoblot analysis with the indicated antibodies. HDAC, a nuclear marker and loading control. **g**, Lysates of primary mouse hepatocytes treated for 0–120 min with IL-6 at 50 ng ml^{-1} were subjected to immunoblot analysis using the indicated antibodies. **h**, Microarray analysis was performed using the Illumina MouseWG-6 v2 Expression BeadChip on RNA extracted from wild-type and *villin-gp130*^{Act} small intestinal crypts ($n = 3$ per group). Data were normalized and analysed as described and expression of the indicated genes is shown as fold-induction compared to wild-type crypts. **i**, RNA was extracted from wild-type and *villin-gp130*^{Act} small intestinal organoids, and *Areg* mRNA expression was measured by qRT–PCR ($n = 3$). Results are averages ± s.e.m.; $*P < 0.05$. Scale bars represent 100 μm (**c**) and all data are representative of at least 2–3 independent experiments.



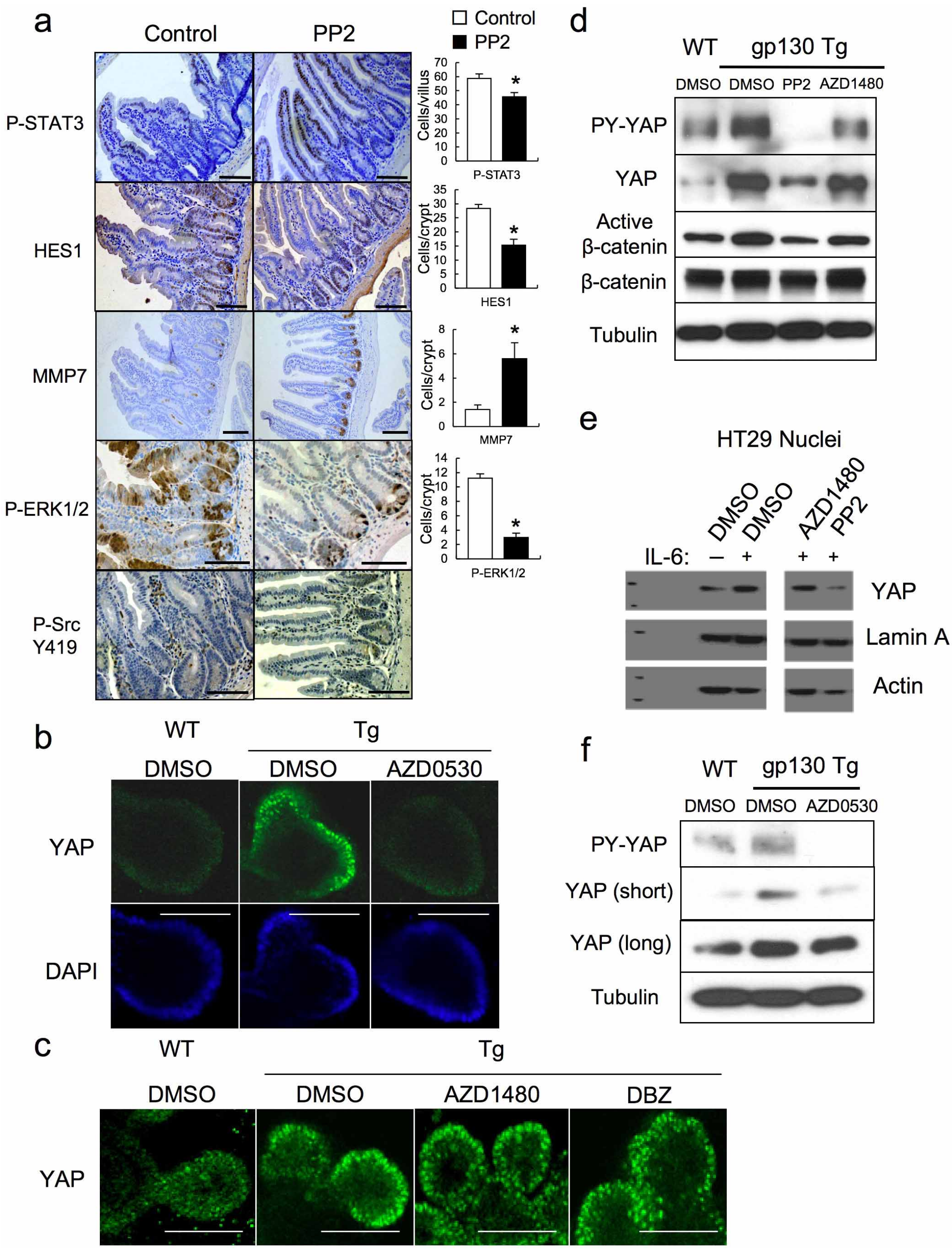
Extended Data Figure 4 | Aberrant intestinal differentiation in *gp130*^{Act} mice depends on Notch and YAP but not on STAT3. a, MMP7 staining of paraffin-embedded sections of control and DBZ-treated (10 μmol kg^{-1}) *villin-gp130*^{Act} small intestines. Positive cells were enumerated in each crypt ($n = 3$). Data represent averages ± s.d.; $*P < 0.05$. **b**, MMP7 and lysozyme staining of paraffin-embedded sections of *villin-gp130*^{Act} and *villin-gp130*^{Act}*/Yap*$^{\Delta IEC}$ small intestines. Positive cells were enumerated in each crypt ($n = 4$). Data represent averages ± s.d.; $*P < 0.05$. **c**, PAS, Ki67, YAP, pSTAT3, HES1 and MMP7 staining of paraffin-embedded sections of *villin-gp130*^{Act} and *villin-gp130*^{Act}*/Stat3*$^{\Delta IEC}$ small intestines. Positive cells were enumerated in each villus or crypt ($n = 4$). Data represent averages ± s.d.; $*P < 0.05$. Scale bars represent 100 μm (**a-c**) and all data are representative of at least 2–3 independent experiments.



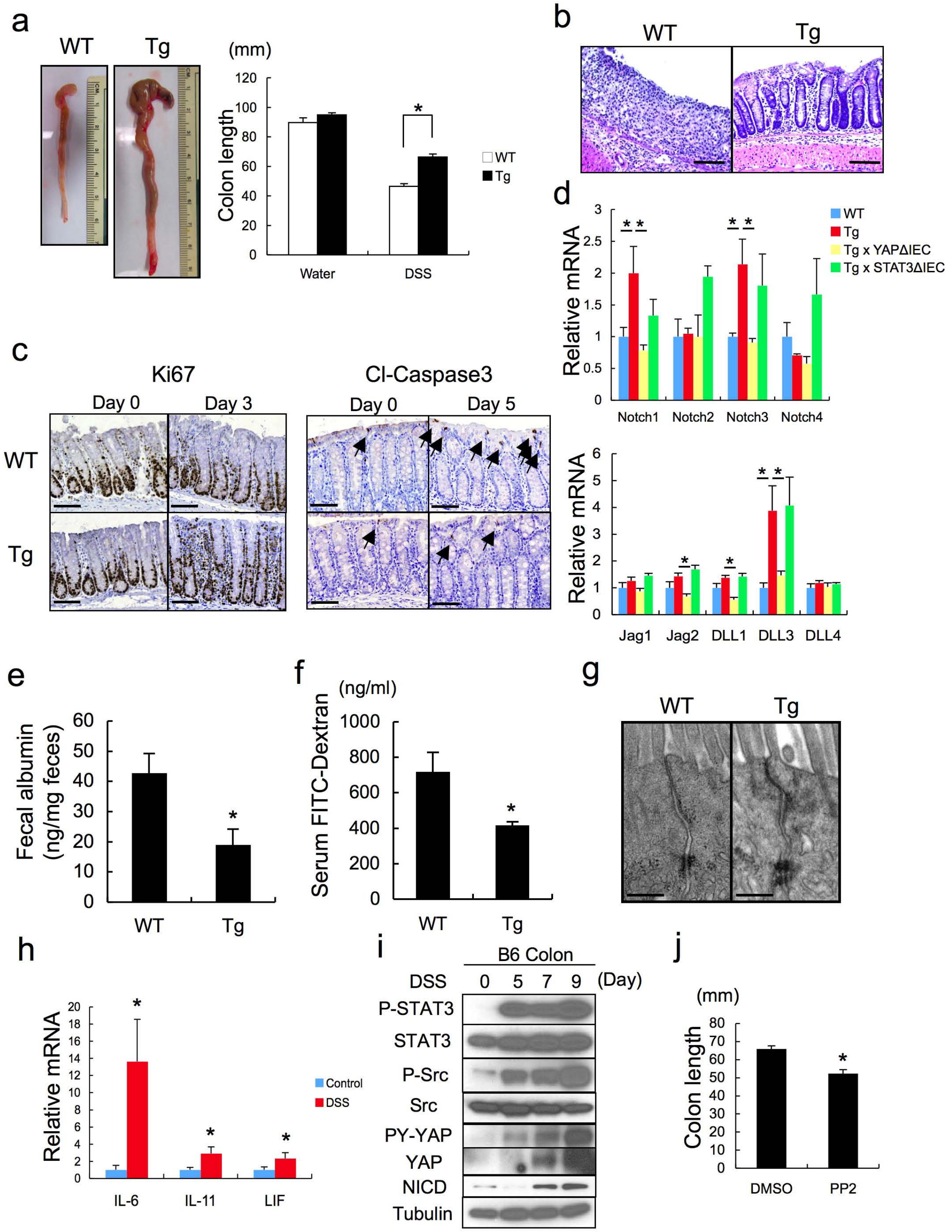
Extended Data Figure 5 | MEK and PI(3)K inhibitors have no effect on aberrant intestinal homeostasis in *gp130^{Act}* mice. **a**, PAS, Ki67, YAP and pERK1/2 staining of paraffin-embedded sections of control and PD0325901-treated (25 mg kg^{-1}) *villin-gp130^{Act}* small intestines. Positive cells were enumerated in each villus or crypt ($n = 3$). Data represent averages ± s.d.; $^*P < 0.05$. **b**, PAS, Ki67, YAP and pS6 staining of paraffin-embedded sections of control and LY294002-treated (100 mg kg^{-1}) *villin-gp130^{Act}* small intestines. Positive cells were enumerated in each villus or crypt ($n = 3$). Data represent averages ± s.d.; $^*P < 0.05$. Scale bars represent 100 μm (**a**, **b**).


a
WT Tg
P-LATS S909
LATS2
GAPDH
b
EV gp130Act gp130SA
P-Src
Src
GAPDH
e
- 10% FBS IL-6 IL-11
P-YAP Y357
YAP
Tubulin
c
Empty gp130 WT gp130Act gp130SA
IP:YAP
P-YAP Y357
YAP
TCL
YAP
Tubulin
d
Empty gp130 WT gp130Act gp130SA
PY-YAP
YAP
P-Src
Src
gp130
Tubulin
h
EV CA-Src
PY-YAP
YAP
CA-Src
Tubulin
f
Con PH
P-Src
YAP
Con PH
PY-YAP
YAP
P-Src
Src
P-STAT3
STAT3
NICD
Tubulin
PH
DMSO PP2
PY-YAP
YAP
P-Src
Src
P-STAT3
STAT3
NICD
Tubulin
g
DMSO PP2 AZD1480
CHX (hr) 0 2 6 0 2 6 0 2 6
Flag-YAP
GAPDH
DMSO AZD0530 SU6656
CHX (hr) 0 2 6 0 2 6 0 2 6
Flag-YAP
Tubulin
i
IP:gp130
Input
– IL-6
150 100 75 50 75 50 37
gp130
Yes
Src
Actin
IP: Src
Input
– IL-6
150 100 75 50 37
gp130
YAP
Src
Actin
j
gp130Act - + - -
gp130Act (Δ771-811) - - + -
gp130Act (Δ812-827) - - - +
Src + + + +
IP:Flag
Src
Flag-gp130
Input
Src
Flag-gp130
k
EV gp130Act gp130Act (Δ771-811) gp130Act (Δ812-827)
YAP
NICD
STAT3
HDAC1
Nuclei

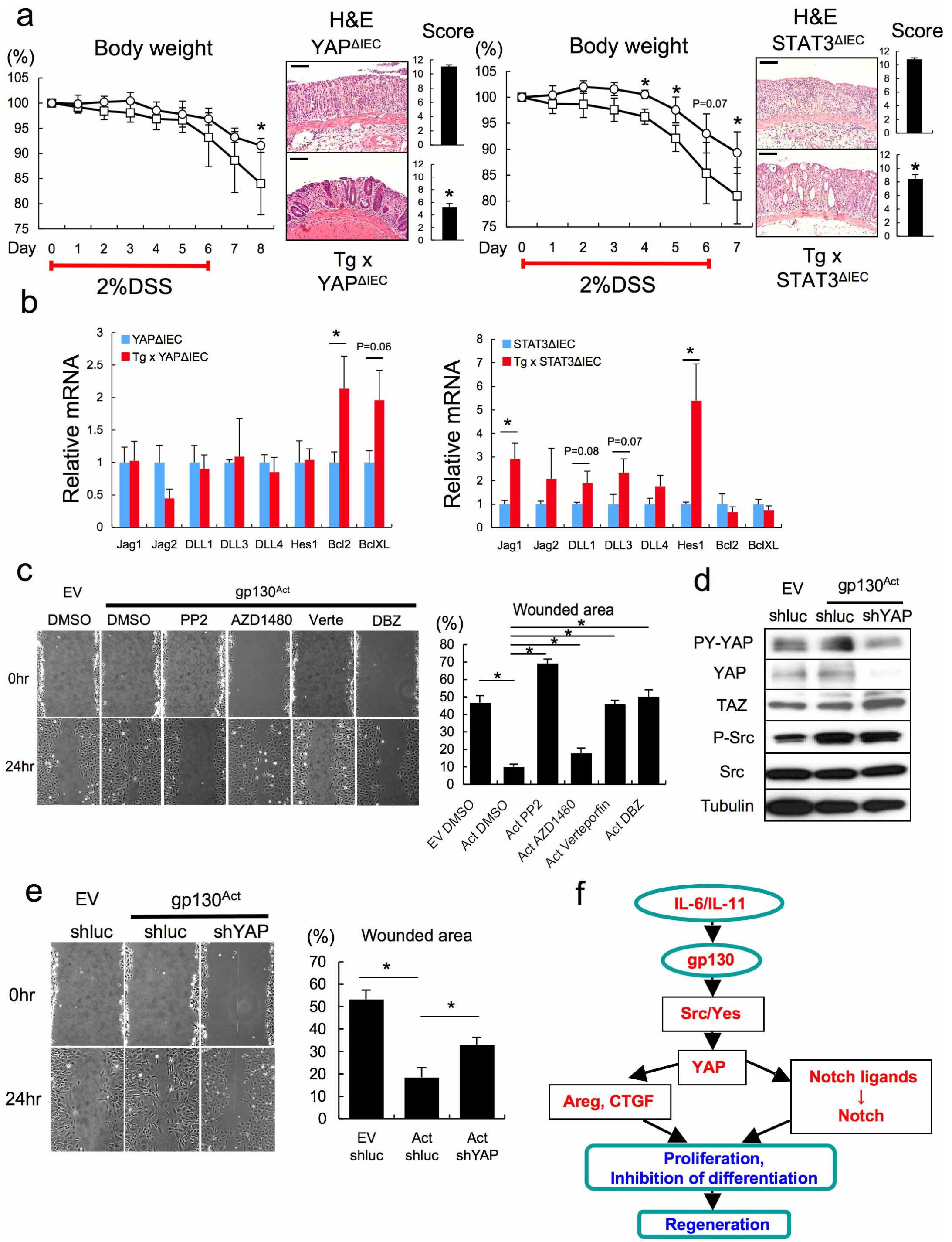
Extended Data Figure 6 | gp130 activates YAP via a Hippo-independent but tyrosine phosphorylation-dependent mechanism, and gp130 interacts with Src and Yes. **a**, Lysates of wild-type and *villin-gp130^{Act}* jejuna, which are the same as the ones in Fig. 2a, were analysed for expression and phosphorylation of the indicated proteins. **b**, Lysates of HT29 colon cancer cells transfected with either empty vector (EV), active gp130 (gp130^{Act}), or superactive gp130 (gp130^{SA}) were subjected to immunoblot analysis with pSrc (Y419), total Src and GAPDH antibodies. GAPDH, a loading control. **c**, Lysates of HCA7 colon cancer cells infected with EV, wild-type gp130, gp130^{Act}, or gp130^{SA} lentiviruses were immunoprecipitated with anti-YAP antibody and blotted with the indicated antibodies. **d**, Lysates of HT29 colon cancer cells infected with EV, wild-type gp130, gp130^{Act}, or gp130^{SA} lentiviruses were immunoblot analysed for expression and phosphorylation of the indicated proteins. **e**, Serum-starved HCT116 cells were stimulated with 10% FBS, IL-6 (100 ng ml^{-1}), or IL-11 (100 ng ml^{-1}) for 30 min. Total cell lysates were analysed for expression and phosphorylation of the indicated proteins. **f**, Left: pSrc and YAP staining of livers from untreated wild-type mice (control) and wild-type mice 48 h after partial hepatectomy (PH). Scale bars represent 100 μm. Middle: lysates of livers from control mice and mice 48 h after partial hepatectomy were subjected to immunoblot analysis with the indicated antibodies. Right: lysates of livers from vehicle (DMSO)-treated and PP2-treated mice 48 h after partial hepatectomy were subjected to immunoblot analysis with the indicated antibodies. **g**, Top: HEK293T cells were transfected with plasmids expressing Flag–YAP. Twenty-four hours after transfection, the cells were pre-treated for 1 h with 0.1% DMSO (vehicle control), PP2 (10 μM) or AZD1480 (1 μM) and then were treated with 50 μg ml^{-1} cycloheximide for different time points. Total cell lysates were subjected to immunoblot analysis with the indicated antibodies. Bottom: HEK293T cells were transfected with Flag–YAP as above. Twenty-four hours after transfection, the cells were pre-treated for 1 h with 0.1% DMSO (vehicle control), AZD0530 (10 μM) or SU6656 (10 μM) and then were treated with 50 μg ml^{-1} cycloheximide for different time points. Total cell lysates were analysed as above. **h**, HEK293T cells were transfected with either empty or constitutively active (CA) Src expression vectors. After 48 h, the cells were lysed and expression of the indicated proteins determined by immunoblot analysis. **i**, HT29 cells were collected with or without 10 ng ml^{-1} IL-6 stimulation for 2 h. Lysates were analysed by immunoblot with the indicated antibodies. These are the loading controls for the data shown in Fig. 4e. **j**, HEK293T cells were transfected with expression vectors encoding Src and Flag-tagged gp130^{Act}, Flag-tagged gp130^{Act} (Δ771–811), Flag-tagged gp130^{Act} (Δ812–827) or empty vector. Cells were collected 48 h later. Cell lysates were immunoprecipitated with Flag antibody and analysed by immunoblot with the indicated antibodies. **k**, Nuclei of T84 colon cancer cells infected with EV, gp130^{Act}, gp130^{Act} (Δ771–811) or gp130^{Act} (Δ812–827) lentiviruses were prepared and subjected to immunoblot analysis with the indicated antibodies. HDAC1, a nuclear marker and loading control. All data are representative of at least 2–3 independent experiments.



Extended Data Figure 7 | SFK activity is required for YAP activation. **a**, Transgenic (Tg) mice ($n = 4$ per group) were treated with PP2 (5 mg kg^{-1}) or vehicle once a day for 5 days. Small intestines were isolated, sectioned and stained as indicated. Positive cells were enumerated in each villus or crypt. Data represent averages ± s.d.; $*P < 0.05$. **b**, **c**, Wild-type and *villin-gp130*^{Act} small intestinal organoids were treated with DMSO, AZD0530 (10 μM) (**b**), AZD1480 (1 μM) or DBZ (10 μM) (**c**) for 24 h, stained with YAP antibody and counter stained with DAPI and photographed under a fluorescent microscope. **d**, Wild-type and *villin-gp130*^{Act} small intestinal organoids were treated with DMSO, PP2 (10 μM) and AZD1480 (1 μM) for 24 h. Total cell lysates were subjected to immunoblot analysis with the indicated antibodies. **e**, Serum-starved HT29 cells were pre-treated for 1 h with 0.1% DMSO (vehicle control), AZD1480 (10 μM) or PP2 (20 μM) before IL-6 (10 ng ml^{-1}) stimulation for 24 h. Nuclear extracts of HT29 cells treated without or with IL-6 in the absence or presence of AZD1480 or PP2 were subjected to immunoblot analysis with the indicated antibodies. Lamin A, a nuclear marker; Actin, a loading control. **f**, Wild-type and *villin-gp130*^{Act} small intestinal organoids were treated with DMSO and AZD0530 (10 μM) for 24 h. Total cell lysates were subjected to immunoblot analysis with the indicated antibodies. Scale bars represent 100 μm (**a**–**c**). All data are representative of at least 2–3 independent experiments.



Extended Data Figure 8 | gp130^{Act} confers DSS resistance, induces Notch receptors and ligands and improves barrier function. **a**, Left: representative images of wild-type and *villin-gp130^{Act}* large intestines taken 10 days after 3.0% DSS treatment. Right: colon length of wild-type and *villin-gp130^{Act}* mice before and after DSS treatment (before: $n = 5$ per group, after: $n = 4$ per group). Results are averages ± s.e.m.; $*P < 0.05$. **b**, Representative images of H&E stained paraffin-embedded colon sections prepared 10 days after DSS challenge of wild-type and transgenic mice as described in Fig. 5a. Scale bars: 100 µm. **c**, Ki67 (left) and cleaved-caspase 3 (right) stainings were performed on paraffin-embedded colon sections from wild-type and transgenic mice at day 0 and 3 (Ki67) or 5 (cleaved-caspase 3) after 3.0% DSS treatment. **d**, Lysates of wild-type, *villin-gp130^{Act}*, *villin-gp130^{Act}/Yap$^{\Delta IEC}$* and *villin-gp130^{Act}/Stat3$^{\Delta IEC}$* colons were prepared, RNA was extracted and expression of the indicated mRNA species was analysed by qRT–PCR ($n = 3$ per group). Results are averages ± s.e.m.; $*P < 0.05$. **e**, **f**, Gut barrier function in wild-type and *villin-gp130^{Act}* mice was examined by measurements of fecal albumin (WT, $n = 6$; Tg, $n = 7$) (**e**) and FITC-Dextran translocation to blood 4 h after oral gavage (WT, $n = 5$; Tg, $n = 4$) (**f**). Results are averages ± s.e.m.; $*P < 0.05$. **g**, TEM images of intestinal mucosa epithelial cell–cell junctions in wild-type and *villin-gp130^{Act}* small intestines. **h**, C57BL/6 mice were given regular water or 2.5% DSS for 7 days. Colonic RNA was extracted on day 10, and expression of the indicated genes was analysed by qRT–PCR ($n = 4$). Results are averages ± s.e.m.; $*P < 0.05$. **i**, Wild-type mice were given 2.5% DSS. Colonic lysates were prepared when indicated and immunoblot analysed for protein expression and phosphorylation. **j**, Colon length of control and PP2-injected C57BL/6 mice after DSS treatment ($n = 6$ per group). Results are averages ± s.e.m.; $*P < 0.05$. Scale bars represent 100 µm (**b**, **c**) and 500 nm (**g**).



Extended Data Figure 9 | Enhanced mucosal regeneration in *gp130*^{Act} mice depends on YAP and STAT3 but the two effectors control different genes, and YAP is required for *in vitro* scratch closure. **a**, Left: body weight curves of DSS-treated *Yap*$^{\Delta IEC}$ (squares, $n = 6$) and *villin-gp130*^{Act}/*Yap*$^{\Delta IEC}$ (circles, $n = 4$) mice. Results are averages ± s.d.; $*P < 0.05$. Colon mucosal histology of *Yap*$^{\Delta IEC}$ (squares, $n = 6$) and *villin-gp130*^{Act}/*Yap*$^{\Delta IEC}$ (circles, $n = 4$) mice was examined by H&E staining and scored 9 days after 2.0% DSS challenge by an observer blinded to the mouse genotype. Results are averages ± s.e.m.; $*P < 0.05$. Right: body weight curves of DSS-treated *Stat3*$^{\Delta IEC}$ (squares) and *villin-gp130*^{Act}/*Stat3*$^{\Delta IEC}$ (circles) mice ($n = 4$ per group). Results are averages ± s.d.; $*P < 0.05$. Mucosal histology of *Stat3*$^{\Delta IEC}$ and *villin-gp130*^{Act}/*Stat3*$^{\Delta IEC}$ mice ($n = 4$ per group) was examined and scored 8 days after 2.0% DSS challenge as above. Results are averages ± s.e.m.; $*P < 0.05$. **b**, RNA was extracted from *Yap*$^{\Delta IEC}$ and *villin-gp130*^{Act}/*Yap*$^{\Delta IEC}$ ($n = 3$ per group) or *Stat3*$^{\Delta IEC}$ and *villin-gp130*^{Act}/*Stat3*$^{\Delta IEC}$ ($n = 4$ per group) colons, and expression of the indicated mRNA species was measured by qRT–PCR. Results are averages ± s.e.m.; $*P < 0.05$. **c**, IEC6 rat intestinal epithelial cells infected with EV or gp130^{Act} lentiviruses were grown to confluence and starved overnight, and the monolayers were wounded by scratching and treated with DMSO, PP2 (10 μM), AZD1480 (1 μM), verteporfin (1 μg ml^{-1}) or DBZ (10 μM). The per cent wounded area was calculated by measuring wound closure over time (0 and 24 h). Results are averages ± s.e.m.; $*P < 0.05$ ($n = 5$). **d**, Total cell lysates of IEC6 cells infected with EV + shluc (control), gp130^{Act} + shluc or gp130^{Act} + shYAP lentiviruses were prepared and subjected to immunoblot analysis with the indicated antibodies. **e**, IEC6 cells infected with EV + shluc (control), gp130^{Act} + shluc or gp130^{Act} + shYAP lentiviruses were grown to confluence and starved overnight, and the monolayers were wounded by scratching. The per cent wounded area was calculated by measuring wound closure over time (0 and 24 h) ($n = 5$). Results are averages ± s.e.m.; $*P < 0.05$. **f**, Schematic representation of the gp130–SFK–YAP–Notch pathway and its function in the injured intestinal epithelium. Scale bars represent 100 μm (**a**).

Extended Data Table 1 | Elevated P-Src and YAP expression in colonic biopsies from CD patients

	Healthy control	CD patient
DN or SP	11	7
DP	0	11

$P<0.05$

DN or SP: Double negative (DN) staining for P-Src and YAP or single positive (SP) staining of either P-Src or YAP.
DP: Double positive (DP) staining for both P-Src and YAP.

Characteristics of Controls and Patients with Crohn's Disease

	Crohn's disease N = 18	Controls N = 11
Age (years +/- S.D.)	53.0 +/- 12.2	39.6 +/- 13.2
Disease duration (years +/- S.D.)	14.7 +/- 13.1	NA
Disease phenotype		NA
Ileocolonic disease	72%	
Colonic disease	28%	
TNF antagonist use	17%	0%
Immune modulator use (azathioprine, 6-mercaptopurine)	17%	0%
5-Aminosalicylate use	56%	0%
Steroid use	5%	0%
Histopathology		
Active chronic colitis	78%	0%
Inactive chronic colitis	22%	0%

N = number of patients, S.D. = standard deviation, NA = not applicable

Shown are the results from pSrc and YAP immunohistochemical analyses, some of whose images are shown in Fig. 4d. Patient characteristics are included.

LETTER

doi:10.1038/nature14124

The double-degenerate, super-Chandrasekhar nucleus of the planetary nebula Henize 2-428

M. Santander-García[1,2], P. Rodríguez-Gil[3,4], R. L. M. Corradi[3,4], D. Jones[3,4], B. Miszalski[5,6], H. M. J. Boffin[7], M. M. Rubio-Díez[8] & M. M. Kotze[5]

The planetary nebula stage is the ultimate fate of stars with masses one to eight times that of the Sun ($M_\odot$). The origin of their complex morphologies is poorly understood[1], although several mechanisms involving binary interaction have been proposed[2,3]. In close binary systems, the orbital separation is short enough for the primary star to overfill its Roche lobe as the star expands during the asymptotic giant branch phase. The excess gas eventually forms a common envelope surrounding both stars. Drag forces then result in the envelope being ejected into a bipolar planetary nebula whose equator is coincident with the orbital plane of the system. Systems in which both stars have ejected their envelopes and are evolving towards the white dwarf stage are said to be double degenerate. Here we report that Henize 2-428 has a double-degenerate core with a combined mass of $\sim$1.76$M_\odot$, which is above the Chandrasekhar limit (the maximum mass of a stable white dwarf) of 1.4$M_\odot$. This, together with its short orbital period (4.2 hours), suggests that the system should merge in 700 million years, triggering a type Ia supernova event. This supports the hypothesis of the double-degenerate, super-Chandrasekhar evolutionary pathway for the formation of type Ia supernovae[4].

The hypothesis of binarity as being essential to producing bipolar planetary nebulae has been progressively gaining ground, particularly with the recent discovery of many new close binary planetary nebula nuclei[5–8]. Here we have undertaken a campaign of photometric monitoring of planetary nebula central stars that show features such as rings and/or jets that are characteristic of central star binarity[9]. One of them, Henize 2-428 (PN G049.4+02.4), has one of the shortest orbital periods found for this class of object. This bipolar nebula consists of two open lobes emerging from a ring-shaped waist inclined at about 68° to the plane of the sky (as deduced from the elliptical shape of the ring in images of the nebula; see Fig. 1). The kinematic age of the bipolar lobes and ring implies that Henize 2-428 is an evolved planetary nebula. However, the presence of a high-density nebular core indicates that strong mass loss and/or mass transfer is occurring close to its central star, which in turn suggests that the nebula may host a binary core[10].

The periodic light curves of Henize 2-428 in the Sloan i-band and Johnson B-band (with effective wavelengths 0.78 μm and 0.44 μm, respectively) demonstrate its binary nature (Fig. 2). They show two broad minima indicative of ellipsoidal modulation due to tidal distortion of one or both components. The period analysis using Schwarzenberg-Czerny's[11] analysis of variance (AOV) method on the photometric data set with the greatest time coverage reveals that the orbital period is 0.1758 ± 0.0005 day. Spectroscopic observations with the FOcal Reducer/low dispersion Spectrograph 2 (FORS2) on the Very Large Telescope (VLT; Fig. 3) show the absorption lines to be double and varying with the orbital phase. The sinusoidal Doppler shifts of the double He II 541.2 nm absorption feature are finally revealed by a series of Optical System for Imaging and low-Intermediate-Resolution Integrated Spectroscopy (OSIRIS) spectra obtained along a full orbit (Fig. 4) on the Gran Telescopio Canarias (GTC). Gaussian fitting of the absorption-line profiles followed by sinusoidal fitting to the data 'folded' (that is, the time of the data is converted into orbital phase) on the orbital period indicates that both radial velocity amplitudes are identical, with values of 206 ± 8 km s^{-1} and 206 ± 12 km s^{-1}, respectively. The very similar depths of the light-curve minima, intensities of the He II 541.2 nm stellar absorption lines, and velocity amplitudes indicate that the two stars have nearly identical masses and effective temperatures. The latter can be constrained by the presence of the photo-ionized nebula and by the lack of He II emission lines in the spectrum of the nebula. The effective temperatures of the stars have therefore been kept between 20,000 K and 40,000 K in our modelling.

The combined analysis of the light curves and radial-velocity curves[12,13] yields the orbital parameters of the system (Table 1). The result is a double-degenerate binary in 'overcontact', that is, both stars are overfilling their Roche lobes. The total luminosity of the system, 845$L_\odot$, where $L_\odot$ is the solar luminosity, is compatible with the 690$L_\odot$ reported in the literature[14].

The inclination of the orbital plane is confined to a narrow range between 63.4° and 66.1° (see Table 1), close to the $\sim$68°-inclined equatorial ring of the nebula, further strengthening the binary hypothesis. It is noteworthy that the nebular and orbital axes of every planetary nebula in which both values are known are coincident within a few degrees. These are A41, A63, A65, HaTr4, NGC 6337, NGC 6778, and Sp1 (ref. 15). This is convincing evidence for the scenario of binaries forming planetary nebulae (see above).

The most striking result is the total mass of the system, which amounts to 1.76$M_\odot$. This is the only reported double-degenerate system whose total mass is definitely above the Chandrasekhar limit[6], even when considering the lowest mass allowed within the confidence range (1.5$M_\odot$).

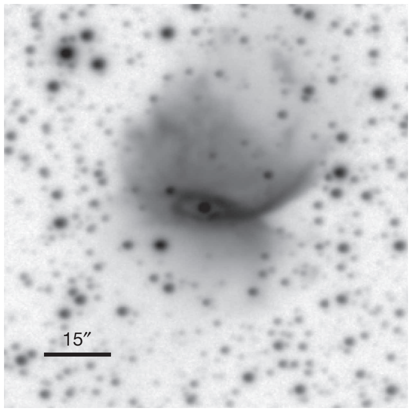


Figure 1 | Close-up view of the bipolar planetary nebula Henize 2-428. This 2-h exposure in Hα 656.3 nm was observed with the INT/WFC. North is up and East is to the left.

[1]Observatorio Astronómico Nacional, Apartado de Correos 112, E-28803, Alcalá de Henares, Spain. [2]Instituto de Ciencia de Materiales de Madrid (CSIC), Sor Juana Inés de la Cruz, 3, E-28049 Madrid, Spain. [3]Instituto de Astrofísica de Canarias, E-38200 La Laguna, Tenerife, Spain. [4]Departamento de Astrofísica, Universidad de La Laguna, E-38205 La Laguna, Tenerife, Spain. [5]South African Astronomical Observatory, PO Box 9, Observatory 7935, South Africa. [6]Southern African Large Telescope Foundation, PO Box 9, Observatory 7935, South Africa. [7]European Southern Observatory, Alonso de Córdova 3107, 19001 Casilla, Santiago, Chile. [8]Centro de Astrobiología, CSIC-INTA, Carretera de Torrejón a Ajalvir, km 4, E-28850 Torrejón de Ardoz, Spain.

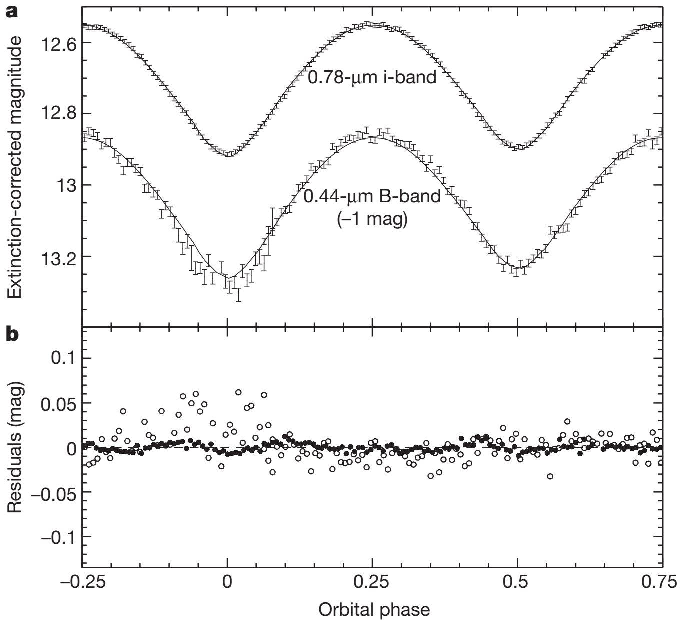


Figure 2 | Light-curve measurements and model. **a**, Light curves of Henize 2-428 in the Johnson B-band and Sloan i-band filters (0.44 µm and 0.78 µm, respectively) and model, along with their respective residuals (**b**). The B-band data have been shifted up by one magnitude for display purposes. The data are shown here folded on the orbital period of the system—0.1758 day (or 4.2 h)—along with the model (solid line). Error bars represent 1σ formal measurement errors.

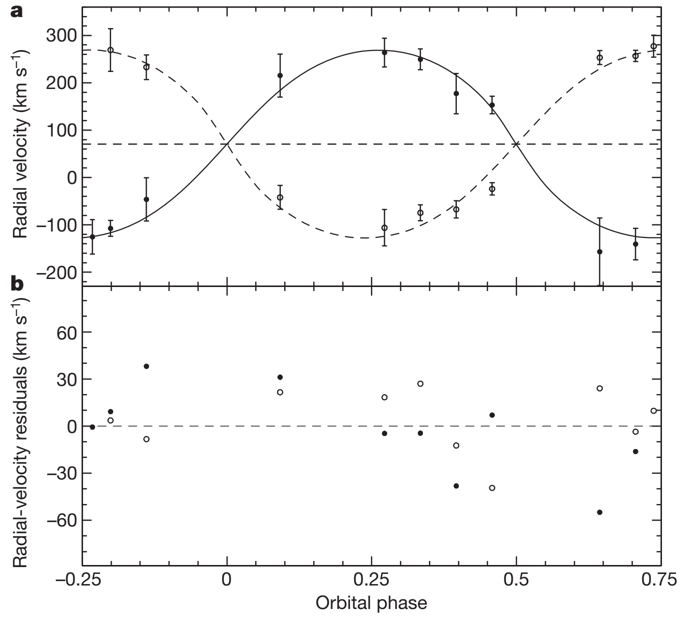


Figure 4 | Radial-velocity measurements and orbit solution. **a**, Radial-velocity curves of the central stars of Henize 2-428 obtained with GTC/OSIRIS on 11 August 2013, and the model, along with their respective residuals (**b**). The data have been folded on the 4.2-h period determined in the text. The primary star is depicted by black points and the secondary star by white ones, and the dashed horizontal line represents the systemic velocity. Error bars represent 1σ formal measurement errors.

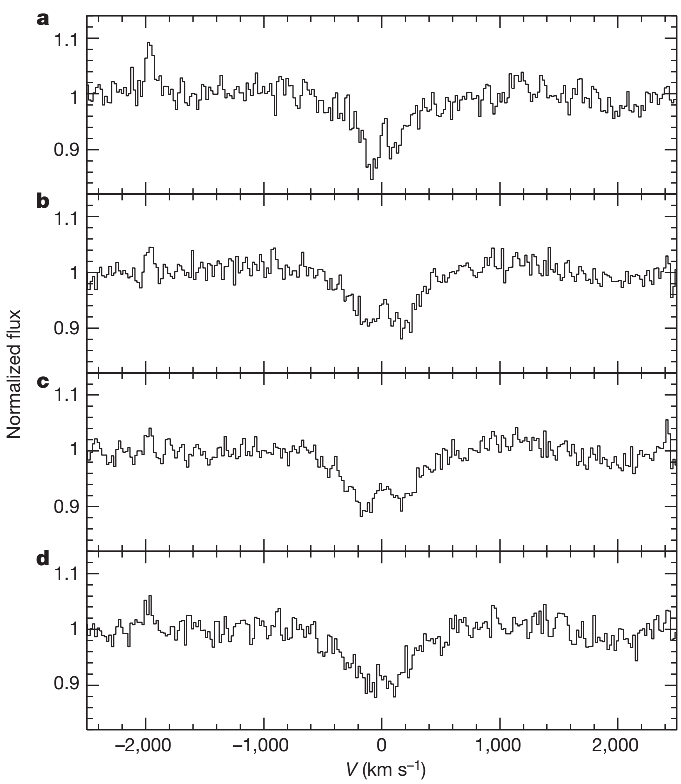


Figure 3 | Time evolution of the spectrum profile of Henize 2-428. The double He II 541.2 nm absorption lines show clear Doppler shifts in the VLT spectra. The flux is normalized with respect to the continuum. Velocities with respect to the He II 541.2 nm rest wavelength are displayed on the x axis. The top spectrum (**a**) corresponds to the night of 19 June 2010, while the three remaining, consecutive spectra were taken on 8 July 2012 and are chronologically ordered from top to bottom (**b**–**d**).

Moreover, with a mass ratio (the mass of the secondary divided by the mass of the primary) of 1, both stars are quite massive compared to average post-asymptotic giant branch (AGB) stars, and their luminosity and effective temperatures do not match the theoretical post-AGB cooling tracks[17]. The explanation for this discrepancy must be sought in their binary evolution. The system seems to have undergone first a phase of mass transfer via wind or stable Roche lobe overflow, and then the formation of a common envelope. This is likely, because for two oversized pre-white-dwarf stars with radius $R = 0.68 - 0.7R_\odot$, where $R_\odot$ is the radius of the Sun, still to be hot, the two events (mass transfer and formation of the common envelope) must have happened fast and consecutively. In addition, for the system to have a current mass ratio close

Table 1 | Orbital and physical parameters of the nucleus of Henize 2-428

Parameter	Model
Orbital period, P (day)	0.1758 ± 0.0005
Epoch of minimum light in the i-band (HJD)	$2456507.53797 \pm 0.00034$
Epoch of minimum light in the B-band (HJD)	$2456484.60636 \pm 0.00077$
Epoch of zero phase in radial velocity data (HJD)	2456516.50377 ± 0.026
Mass ratio, $q \equiv M_2/M1$	1
Orbital separation, a ($R_\odot$)	1.59 ± 0.06
Inclination, i (°)	64.7 ± 1.4
Centre-of-mass redshift, γ (km s^{-1})	71 ± 7
Primary mass, M_1 ($M_\odot$)	0.88 ± 0.13
Secondary mass, M_2 ($M_\odot$)	0.88 ± 0.13
Primary radius, R_1 (pole, $R_\odot$)	0.68 ± 0.04
Secondary radius, R_2 (pole, $R_\odot$)	0.68 ± 0.04
Primary surface potential[12], Ω_1	3.50 ± 0.06
Secondary surface potential[12], Ω_2	3.50 ± 0.06
Primary temperature, T_1 ($\times 10^3$ K)	32.4 ± 5.2
Secondary temperature, T_2 ($\times 10^3$ K)	30.9 ± 5.2
Distance to the system, d (kpc)	1.4 ± 0.4

Orbital and physical parameters of the nucleus of Henize 2-428. They were determined by modelling the INT/WFC Sloan i-band and the SAAO 1 m telescope/SHOC (Sutherland High-speed Optical Cameras) Johnson B-band light-curve data, together with the GTC/OSIRIS spectra, all folded on the orbital period. A systematic search of the parameter space was performed on the inclination, orbital separation, centre-of-mass redshift, temperatures and surface potential of both stars until χ^2 was globally minimized. Uncertainties in the parameters represent standard 1σ formal errors. See the Methods section for details of the modelling process.

to 1, the initial mass ratio must have been also very close to 1. Such a case would lead to a stable Roche lobe overflow initially, given the relatively massive cores and judging by stability studies[18]. After the first pre-white-dwarf star was formed (from a quite massive AGB[19]), the mass ratio (AGB/pre-white-dwarf) would then be much larger than 1, as the secondary is still a massive AGB as well, and the system would undergo a common-envelope phase. Having an initial mass ratio close to 1 would also explain the final, similar luminosities and effective temperatures.

An alternative model in which one or both stars are non-degenerate helium stars deserves further discussion. A $0.88M_{\odot}$ helium star approaching core helium exhaustion[20], produced by the hydrogen shell of a massive star being stripped off during Roche lobe overflow, would have a compatible luminosity of $422L_{\odot}$, although its radius $0.16R_{\odot}$ is still lower than the $0.68R_{\odot}$ radius found in this work, and the position of the stars of Henize 2-428 in $\log(g\text{–}T_{\rm eff})$ (gravity versus effective temperature) diagrams does not match the locus of helium stars in binary systems[21] in the same way as it does not match the aforementioned post-AGB cooling tracks. This alternative model is, however, very unlikely, considering that, first, the computed He/H and the supersolar N/O abundance values[10] indicate a composition more typical of a processed, AGB envelope than a red giant branch envelope; and, second, gas densities in the extended equatorial ring are around 1,000 cm^{-3}, implying recombination times of the order of a few tens of years for H, and a few years for N^{+}, O^{+}, and O^{++}. Given the nebula age of the order of a few thousand years[10], this excludes the possibility that the nebula has a fossil photoionization. In any case, more robust theoretical models should help us understand the peculiarities of this intriguing system.

The stars are massive and close enough to merge in a Hubble time[22] (~700 million years), thus making Henize 2-428 a strong type Ia supernova candidate. Such a massive double-degenerate system may well explain supernova remnants like SNR 0509-67.5, in which the single-degenerate scenario can be ruled out[23], and confirms the plausibility of double-degenerate scenarios as type Ia supernova progenitors.

The thermal timescale of the stars is ~80,000 years, much less than the merging time of the system. If the radius of each star contracts on the thermal timescale, then the system will relatively quickly become detached and the orbital period will shrink accordingly. If this is the case, long-term monitoring of this system might reveal small changes in the orbital period over the course of the next decades.

In addition to Henize 2-428, only half a dozen double-degenerate binary systems are known so far: V458 Vul[24]; SBS 1150+599A, the nucleus of TS 01[16]; MT Ser, the nucleus of Abell 41[25], NGC 6026[26], and Fg 1[6]. However, estimates indicate they could comprise up to 25% of the total planetary nebulae hosting a close-binary central star[27]. This figure could be even larger given the strong selection effects against double degenerates in photometric surveys—in fact, most double-degenerate systems may only be identified via radial velocity studies[6]. A larger sample of double-degenerate central stars of planetary nebulae would help us to assess their contribution to the total number of planetary nebulae or type Ia supernova progenitors, and also to understand the role of common-envelope and/or contact-binarity in the evolution and shaping of planetary nebulae and in the production of type Ia supernova events.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 September; accepted 28 November 2014.
Published online 9 February 2015.

1. Balick, B. & Frank, A. Shapes and shaping of planetary nebulae. *Annu. Rev. Astron. Astrophys.* **40,** 439–486 (2002).
2. De Marco, O. The origin and shaping of planetary nebulae: putting the binary hypothesis to the test. *Publ. Astron. Soc. Pacif.* **121,** 316–342 (2009).
3. García-Arredondo, F. & Frank, A. Collimated outflow formation via binary stars: three-dimensional simulations of asymptotic giant branch wind and disk wind interactions. *Astrophys. J.* **600,** 992–1003 (2004).
4. Howell, D. A. *et al.* The type Ia supernova SNLS-03D3bb from a super-Chandrasekhar-mass white dwarf star. *Nature* **443,** 308–311 (2006).
5. Boffin, H. M. J. & Miszalski, B. in *Evolution of Compact Binaries* (eds Schmidtobreick, L., Schreiber, M. R. & Tappert, C.) *Astron. Soc. Pacif. Conf. Ser.* **447,** 159–164 (2011).
6. Boffin, H. M. J. *et al.* An interacting binary system powers precessing outflows of an evolved star. *Science* **338,** 773–775 (2012).
7. Jones, D. *et al.* The post-common-envelope, binary central star of the planetary nebula Hen 2–11. *Astron. Astrophys.* **562,** 89–97 (2014).
8. Corradi, R. L. M. *et al.* The planetary nebula IPHASXJ211420.0+434136 (Ou5): insights into common-envelope dynamical and chemical evolution. *Mon. Not. R. Astron. Soc.* **441,** 2799–2808 (2014).
9. Miszalski, B., Acker, A., Parker, Q. A. & Moffat, A. F. J. Binary planetary nebulae nuclei towards the Galactic bulge. II. A penchant for bipolarity and low-ionisation structures. *Astron. Astrophys.* **505,** 249–263 (2009b).
10. Rodríguez, M., Corradi, R. L. M. & Mampaso, A. Evidence for binarity in the bipolar planetary nebulae He 2–428 and M 1–91. *Astron. Astrophys.* **377,** 1042–1055 (2001).
11. Schwarzenberg-Czerny, A. Fast and statistically optimal period search in uneven sampled observations. *Astrophys. J. Lett.* **460,** L107–L110 (1996).
12. Wilson, R. E. Eccentric orbit generalization and simultaneous solution of binary star light and velocity curves. *Astrophys. J.* **234,** 1054–1066 (1979).
13. Prsa, A., Matijevic, G., Latkovic, O., Vilardell, F. & Wils, P. *PHOEBE: PHysics Of Eclipsing BinariEs* (Astrophysics Source Code Library, 2011).
14. Stanghellini, L., Villaver, E., Manchado, A. & Guerrero, M. A. The correlations between planetary nebula morphology and central star evolution: analysis of the northern galactic sample. *Astrophys. J.* **576,** 285–293 (2002).
15. Jones, D., Santander-Garcia, M., Boffin, H. M. J., Miszalski, B. & Corradi, R. L. M. The morpho-kinematics of planetary nebulae with binary central stars. In *Proc. Asymmetrical Planetary Nebulae VI Conf. (4–8 November 2013)* (eds Morisset, C., Delgado-Inglada, G. & Torres-Peimbert, S.) 43 (2014).
16. Tovmassian, G. *et al.* The double-degenerate nucleus of the planetary nebula TS 01: a close binary evolution showcase. *Astrophys. J.* **714,** 178–193 (2010).
17. Bloecker, T. Stellar evolution of low- and intermediate-mass stars. II. Post-AGB evolution. *Astron. Astrophys.* **299,** 755–769 (1995).
18. Chen, X. & Han, Z. Mass transfer from a giant star to a main-sequence companion and its contribution to long-orbital-period blue stragglers. *Mon. Not. R. Astron. Soc.* **387,** 1416–1430 (2008).
19. Weidemann, V. Revision of the initial-to-final mass relation. *Astron. Astrophys.* **363,** 647–656 (2000).
20. Paczyński, B. Evolutionary processes in close binary systems. *Annu. Rev. Astron. Astrophys.* **9,** 183–208 (1971).
21. Stroeer, A. *et al.* Hot subdwarfs from the ESO supernova Ia progenitor survey. II. Atmospheric parameters of subdwarf O stars. *Astron. Astrophys.* **462,** 269–280 (2007).
22. Shapiro, S. L. & Teukolsky, S. A. *Black Holes, White Dwarfs, and Neutron Stars: the Physics of Compact Objects* (Wiley-Interscience, 1983).
23. Schaefer, B. E. & Pagnotta, A. An absence of ex-companion stars in the type Ia supernova remnant SNR 0509–67.5. *Nature* **481,** 164–166 (2012).
24. Rodríguez-Gil, P. *et al.* The orbital period of V458 Vulpeculae, a post-double common-envelope nova. *Mon. Not. R. Astron. Soc.* **407,** L21–L25 (2010).
25. Bruch, A., Vaz, L. P. R. & Diaz, M. P. An analysis of the light curve of the post common envelope binary MT Serpentis. *Astron. Astrophys.* **377,** 898–910 (2001).
26. Hillwig, T. C., Bond, H. E., Afşar, M. & De Marco, O. Binary central stars of planetary nebulae discovered through photometric variability. II. Modeling the central stars of NGC 6026 and NGC 6337. *Astron. J.* **140,** 319–327 (2010).
27. Hillwig, T. C. The current status of our understanding of the close binary central stars of planetary nebulae. In *18th European White Dwarf Workshop* (eds Krzesiński, J., Stachowski, G., Moskalik, P. & Bajan, K.) *Astron. Soc. Pacif. Conf. Ser.* **469,** 277 (2013).

Acknowledgements This work is based on observations made with the 1 m SAAO (South Africa Astronomical Observatory), the 1.2 m Mercator, the 2.5 m INT (Isaac Newton Telescope), the 4.2 m WHT (William Herschel Telescope), the 8.2 m VLT and the 10.4 m GTC telescopes. We are grateful to T. Marsh for the use of the PAMELA and MOLLY codes, to T. Hillwig, O. Pols and J. Alcolea for their comments and to J. García-Rojas and C. Zurita for the INT/WFC (Wide Field Camera) service observations. This work was partially supported by the Spanish MINECO within grants CSD2009–00038, AYA2012–35330, RYC–2010–05762 and AYA 2012–38700.

Author Contributions M.S.-G., P.R.-G., D.J., M.M.R.-D., H.M.J.B. and M.M.K. conducted the observations at the various telescopes. M.S.-G., P.R.-G., D.J. and M.M.K. reduced the data. M.S.-G. performed the light-curve and radial-velocity-curve modelling, and wrote the paper. All authors discussed the results and implications and commented on the manuscript at all stages.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S.-G. (m.santander@oan.es).

METHODS

Information on the observational data. Four short series of i-band time-resolved photometry of Henize 2-428 were taken with MEROPE[28] (the MERcator Optical Photometric ImagEr) on the Mercator telescope[29] on La Palma on 28 and 30 August 2009. They showed a photometric variability as large as ~0.36 mag between series, so the system was monitored for a single, 4-h interval on the night of 2 September 2009, and an orbital period was determined. Another similar 4-h time series was covered in the Johnson B-band with the SAAO 1 m telescope/SHOC on July 11, 2013, and in the Sloan i-band with the INT/WFC on 2 August 2013.

A single 20-min spectrum with VLT/FORS2 was secured in the blue range on 19 June 2010, under program ID 085.D-0629(A), and three additional 15-min spectra with the same configuration on the night of 8 July 2012, with program ID 089.D-0453(A). The '1200g' grism with a slit width of 0.7 arcsec was used in all cases. The spectra covered the range 409–556 nm. The resulting effective resolution was 0.8 Å. These four spectra taken at different times during an orbit clearly showed radial-velocity variations from both stars and established the need for and feasibility of a systematic radial-velocity study (see Fig. 3). This was carried out with GTC/OSIRIS on 11 August 2013. The object was monitored for 3.8 h with a slit width of 0.6 arcsec at the parallactic angle. The 'R2000B' grating was used, resulting in a wavelength coverage from 396 nm to 569.5 nm, and the effective resolution was 1.9 Å. The spectra were binned once in the spatial direction.

Data modelling. The magnitude values form the INT/WFC Sloan i-band and the SAAO 1 m telescope/SHOC Johnson B-band light-curve data were corrected for an extinction of $A_v = 2.96 \pm 0.34$, recomputed from an available value[10] using a different extinction law[30]. We performed a period analysis using Schwarzenberg-Czerny's[11] AOV method on the photometric data set with the greatest time coverage (from Mercator/MEROPE covering 6.25 h along three nights, not shown here). The AOV periodogram shows the strongest peak at ~11.379 cycles per day, which would correspond to a period of 0.0879 day (or 2.1 h). The orbital period of the system is, however, twice as long, 0.1758 ± 0.0005 day (4.2 h), as indicated by the ellipsoidal modulation of the light curve, with the two minima showing similar but different depths.

The ephemerides of the light curves indicated in Table 1 were computed from their respective light-curve data. Owing to the asymmetry in the radial-velocity curves, the corresponding ephemeris of the zero phase was computed from the ephemeris of the Sloan i-band light-curve data, taken nine nights before, and therefore the associated, accumulated error is much larger.

The orbital and physical parameters of the nucleus of Henize 2-428 were determined by modelling the light curves, together with the radial-velocity curves, all folded on the orbital period, 0.1758 day. The mass ratio q was fixed to 1, as suggested by the amplitude of the radial-velocity curves. A systematic search of the parameter space was performed on the inclination, orbital separation, centre-of-mass velocity, temperatures and surface potential of both stars until χ^2 was globally minimized. Given the high effective temperature of both stars, an albedo of 1.0 was used for both components. Gravity brightening and square-root limb darkening coefficients were computed according to the temperature and gravity of each component[31,32].

We provide a rough estimate of the distance to Henize 2-428 from the comparison of the model's total luminosity to the dereddened, apparent magnitudes, adopting a bolometric correction according to the stars effective temperatures. The resulting distance is 1.4 ± 0.4 kpc.

B-band data phase determination and alternate model. Observations in the B-band were taken 22 nights before the data used to determine the orbital phase; error accumulation during this time interval amounts to half an orbital period, thus preventing accurate phase determination of the B-band data. Therefore, actually two possibilities were independently considered and modelled; one where the deepest minimum in both light curves occurs at orbital phase 0 (model shown in the paper), and another one where the deepest minimum is offset by half an orbit.

The results of both models are very similar within uncertainties. These include the inclination, which is confined to a narrow range between 62.2° and 63.8°, still a few degrees away from the ~68°-inclined equatorial ring of the nebula. The surface potentials of both stars are slightly different in this model, and the temperature of the primary is around a thousand kelvin lower than that of the secondary. The total mass is slightly larger but similar to the model shown in the paper, 1.84 $M_\odot$, still above the Chandrasekhar limit even when considering uncertainties. The total luminosity of the system is somewhat larger, 1,200$L_\odot$.

There is not enough solid ground to adopt one model over the other.

Code availability. Every code used for this work is third-party and public. In particular, we used PAMELA (now part of the Starlink distribution, available at http://starlink.eao.hawaii.edu/starlink) to reduce the spectra; IRAF (http://iraf.noao.edu) to reduce the images and perform the photometric analysis; MOLLY (http://www2.warwick.ac.uk/fac/sci/physics/research/astro/people/marsh/software/) to analyse the spectra; and PHOEBE (http://phoebe-project.org/1.0/) to model the light and radial-velocity curves.

28. Davignon, G. *et al.* CCD camera and automatic data reduction pipeline for the Mercator telescope on La Palma. In *Ground-based Instrumentation for Astronomy* (eds Moorwood, A. F. M. & Iye, M.) *SPIE Conf. Ser.* **5492,** 871–879 (2004).
29. Raskin, G. *et al.* Mercator and the P7-2000 photometer. In *Ground-based Instrumentation for Astronomy* (eds Moorwood, A. F. M. & Iye, M.) *SPIE Conf. Ser.* **5492,** 830–840 (2004).
30. Fitzpatrick, E. L. Interstellar Extinction in the Milky Way Galaxy. In *Astrophysics of Dust* (eds Witt, A. N., Clayton, G. C. & Draine B. T.), vol. 309 of *Astron. Soc. Pacif. Conf. Ser.* **309,** 33–56 (2004).
31. Claret, A. & Bloemen, S. Gravity and limb-darkening coefficients for the Kepler, CoRoT, Spitzer, uvby, UBVRIJHK, and Sloan photometric systems. *Astron. Astrophys.* **529,** 75–79 (2011).
32. Castelli, F. & Kurucz, R. L. New grids of ATLAS9 model atmospheres. Preprint at http://arxiv.org/abs/astro-ph/0405087 (2004).

LETTER

doi:10.1038/nature14270

State preservation by repetitive error detection in a superconducting quantum circuit

J. Kelly[1]*, R. Barends[1]†*, A. G. Fowler[1,2]†*, A. Megrant[1,3], E. Jeffrey[1]†, T. C. White[1], D. Sank[1]†, J. Y. Mutus[1]†, B. Campbell[1], Yu Chen[1]†, Z. Chen[1], B. Chiaro[1], A. Dunsworth[1], I.-C. Hoi[1], C. Neill[1], P. J. J. O'Malley[1], C. Quintana[1], P. Roushan[1]†, A. Vainsencher[1], J. Wenner[1], A. N. Cleland[1] & John M. Martinis[1]†

Quantum computing becomes viable when a quantum state can be protected from environment-induced error. If quantum bits (qubits) are sufficiently reliable, errors are sparse and quantum error correction (QEC)[1–6] is capable of identifying and correcting them. Adding more qubits improves the preservation of states by guaranteeing that increasingly larger clusters of errors will not cause logical failure—a key requirement for large-scale systems. Using QEC to extend the qubit lifetime remains one of the outstanding experimental challenges in quantum computing. Here we report the protection of classical states from environmental bit-flip errors and demonstrate the suppression of these errors with increasing system size. We use a linear array of nine qubits, which is a natural step towards the two-dimensional surface code QEC scheme[7], and track errors as they occur by repeatedly performing projective quantum non-demolition parity measurements. Relative to a single physical qubit, we reduce the failure rate in retrieving an input state by a factor of 2.7 when using five of our nine qubits and by a factor of 8.5 when using all nine qubits after eight cycles. Additionally, we tomographically verify preservation of the non-classical Greenberger–Horne–Zeilinger state. The successful suppression of environment-induced errors will motivate further research into the many challenges associated with building a large-scale superconducting quantum computer.

The ability to withstand multiple errors during computation is a critical aspect of error correction. We define nth-order fault tolerance to mean that any combination of n errors is tolerable. Previous experiments based on nuclear magnetic resonance[8,9], ion traps[10] and superconducting circuits[11–13] have demonstrated multi-qubit states that are first-order tolerant to one type of error. Recently, experiments with ion traps and superconducting circuits have shown the simultaneous detection of multiple types of errors[14,15]. All of these experiments demonstrate error correction in a single round; however, quantum information must be preserved throughout computation using multiple error-correction cycles. The basics of repeating cycles have been shown in ion traps[16] and superconducting circuits[17]. Until now, it has been an open challenge to combine these elements to make the information stored in a quantum system robust against errors which intrinsically arise from the environment.

The key to detecting errors in quantum information is to perform quantum non-demolition (QND) parity measurements. In the surface code, this is done by arranging qubits in a chequerboard pattern—with data qubits corresponding to the white squares (blue in Fig. 1), and measurement qubits to the black squares (green in Fig. 1)—and using these ancilla measurement qubits to repetitively perform parity measurements to detect bit-flip ($\hat{X}$) and phase-flip ($\hat{Z}$) errors[7,18]. A square chequerboard with $(4n+1)^2$ qubits is nth-order fault tolerant, meaning that at least $n+1$ errors must occur to cause failure in preserving a state if fidelities are above a threshold. With error suppression factor $\Lambda > 1$ and more qubits, failure becomes increasingly unlikely with probability $\varepsilon \propto -1/\Lambda^{n+1}$ (assuming independent errors). The surface code is highly appealing for superconducting quantum circuits as it requires only nearest-neighbour interactions, single and two-qubit gates, and fast repetitive measurements with fidelities above a lenient threshold of approximately 99%. All of this has recently been demonstrated in separate experiments[19,20].

The simplest system demonstrating the basic elements of the surface code is a one-dimensional chain of qubits, as seen in Fig. 1a. It can run the repetition code, a primitive of the surface code, which corrects bit-flip errors on both data and measurement qubits. The device shown in Fig. 1b is a chain of nine qubits, which allows us to experimentally test both first- and second-order fault tolerance. It consists of a superconducting aluminium film on a sapphire substrate, patterned into Xmon transmon qubits[21] with individual control and readout. The qubits are the cross-shaped devices; they are capacitively coupled to their nearest neighbours, controlled with microwave drive and frequency detuning pulses, and measured with a dispersive readout scheme. The device consists of five data qubits and four measurement qubits in an alternating pattern; see Supplementary Information for details.

To detect bit-flips, we determine the parity of adjacent data qubits by measuring the operator $\hat{Z}\hat{Z}$. We do this using an ancilla measurement qubit, and performing single- and two-qubit quantum gates (Fig. 1c). The operator measurement can have two values: $+1$ for states $|00\rangle$ and $|11\rangle$, and -1 for $|01\rangle$ and $|10\rangle$. Therefore, errors can be detected as they occur by repeating this operator measurement and noting changes in the outcome. Importantly, this measurement does not destroy the quantum nature: given input $a|00\rangle + b|11\rangle$ the result will be $+1$ and the quantum state remains, with similar behaviour for other Bell-like superposition states. In the repetition code, simultaneous measurements of these operators enable multiple bit-flip errors to be detected.

We now discuss how bit-flip errors, which can occur on any qubit and at any time, are identified. The quantum circuit of the repetition code is shown in Fig. 2a, for three cycles (in time) and nine qubits. This is the natural extension of the schematic in Fig. 1c, optimized for our hardware (Supplementary Information). Figure 2a illustrates four distinct types of bit-flip errors (stars): measurement error (gold), single-cycle data error (purple), two-cycle data error (red), and a data error after the final cycle (blue). Controlled-NOT (CNOT) gates propagate bit-flip errors on the data qubit to the measurement qubit. Each of these errors is typically detected at two locations if in the interior and at one location if at the boundary; we call these 'detection events'. The error connectivity graph[22] is shown in Fig. 2b, where the grey lines indicate every possible pattern of detection events that can arise from a single error. The last column of values for the $\hat{Z}\hat{Z}$ operators in Fig. 2b are constructed from the data qubit measurements, so that errors between the last cycle and data qubit measurement can be detected (Supplementary Information).

In the absence of errors, there are two possible patterns of sequential measurement results. If a measurement qubit's neighbouring data

[1]Department of Physics, University of California, Santa Barbara, California 93106, USA. [2]Centre for Quantum Computation and Communication Technology, School of Physics, The University of Melbourne, Victoria 3010, Australia. [3]Department of Materials, University of California, Santa Barbara, California 93106, USA. †Present address: Google Inc., Santa Barbara, California 93117, USA.
*These authors contributed equally to this work.

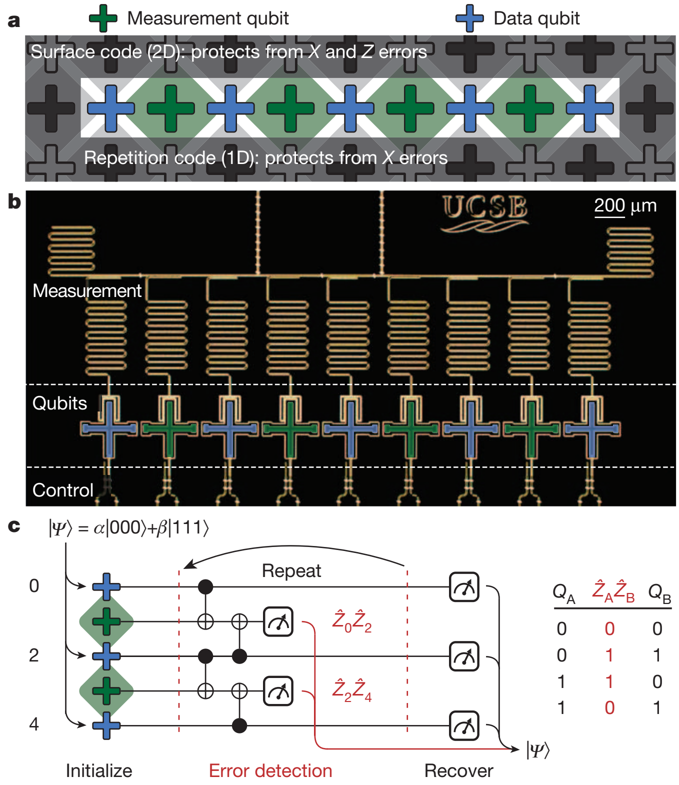


Figure 1 | **Repetition code: device and algorithm.** **a**, The repetition code is a one-dimensional (1D) variant of the surface code, and is able to protect against $\hat{X}$ (bit-flip) errors. The code is implemented using an alternating pattern of data and measurement qubits. **b**, Optical micrograph of the superconducting quantum device, consisting of nine Xmon[21] transmon qubits with individual control and measurement, with a nearest-neighbour coupling scheme. **c**, The repetition code algorithm uses repeated entangling and measurement operations which detect bit-flips, using the parity scheme on the right. Using the output from the measurement qubits during the repetition code for error detection, the initial state can be recovered by removing physical errors in software. Measurement qubits are initialized into the $|0\rangle$ state and need no reinitialization as measurement is QND.

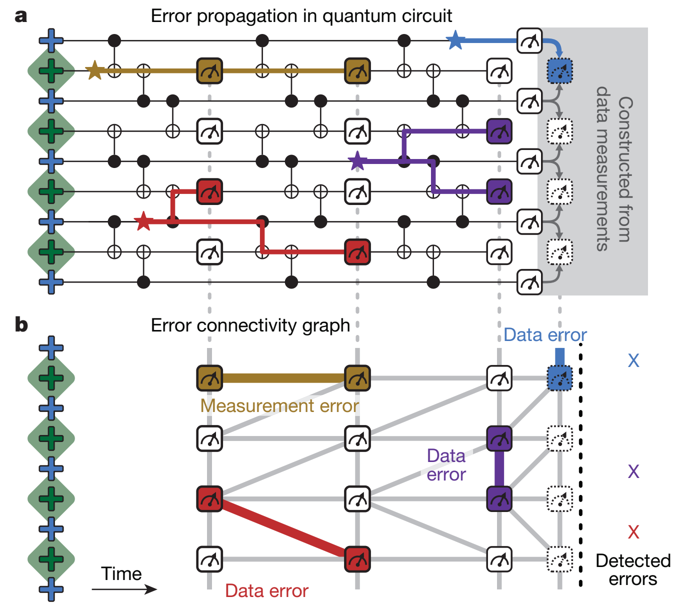


Figure 2 | **Error propagation and identification.** **a**, The quantum circuit for three cycles of the repetition code, and examples of errors. Errors propagate horizontally in time, and vertically through entangling gates. Different errors lead to different detection patterns: an error on a measurement qubit (gold) is detected in two subsequent rounds. Data qubit errors (purple, red, blue) are detected on neighbouring measurement qubits in the same or next cycle. Data errors after the last round (blue) are detected by constructing the final set of $\hat{Z}\hat{Z}$ eigenvalues from the data qubit measurements. **b**, The connectivity graph for the quantum circuit above, showing measurements and possible patterns of detection events (grey), see main text for details. The example detection events and their connections are highlighted, and the corresponding detected errors are shown on the right, which when applied, will recover the input data qubit state.

qubits are in the $|00\rangle$ or $|11\rangle$ state, the measurement qubit will report a string of identical values. If the data qubits are in the $|01\rangle$ or $|10\rangle$ state, the measurement qubit will report alternating values, as measurement is QND. Single data bit-flip errors make the measurement outcomes switch between these two patterns. For example, if the measurement outcomes for three cycles are 0, 0 and 1, this indicates a change from the identical to the alternating pattern in the last measurement, and hence a detection event. Explicitly, with m_t the measurement qubit outcome at cycle t and $\oplus$ the exclusive OR (XOR) operator, for each of the two patterns we have $b_t = m_{t-1} \oplus m_t = 0$ or 1. A detection event at cycle t is then identified when $D_t = b_{t-1} \oplus b_t = 1$.

We use minimum-weight perfect matching[23–25] to decode to physical errors, based on the pattern of detection events and an error model for the system. Intuitively, such matching connects detection events in pairs or to the boundary using the shortest total weighted path length. It is important to note that errors can lead to detection event pairs that span multiple cycles, necessitating the need for multi-round analysis as opposed to round-by-round (see Supplementary Information for details).

To study the ability of our device to preserve quantum states, we initialized the data qubits into a Greenberger–Horne–Zeilinger (GHZ) state, $\left[(|000\rangle + |111\rangle)/\sqrt{2}\right]$, and applied two rounds of the repetition code (Fig. 3). The algorithm is shown in Fig. 3a. Using quantum state tomography we measured the input density matrix ρ and find a GHZ state with fidelity $\mathrm{Tr}(\rho_{\mathrm{ideal}}\rho)$ of 82%, above the threshold of 50% for genuine entanglement[26]. After two repetition code cycles, we use tomography to construct the density matrices for each pattern of detection events. We find a state fidelity of 78% in the case of no detection events, indicating a retention of genuine quantum entanglement. In the case of two detection events, which indicate a likely data qubit error in the first cycle, we find elements away from the ideal positions. By applying the recovery operation in post-processing (a single bit-flip on the blue data qubit) we can restore the state. Energy relaxation, the most likely cause of the detected bit-flip error, induces both bit-flip and phase-flip errors. The bit-flip error is corrected and the diagonal terms are preserved, but any phase-flip error remains uncorrected, reducing the off-diagonal terms and fidelity to 59%. We note that genuine entanglement is preserved. Conditional tomography for every configuration can be found in Supplementary Information.

The data in Fig. 3 clearly show that the one-dimensional repetition code algorithm does not necessarily destroy the quantum nature of the state. It allows for preserving the quantum state in the case of no errors, and correcting bit-flip errors otherwise. This preservation is achieved purely through error detection and classical post-processing, like for the full surface code, avoiding the need for dynamic feedback with quantum gates. For the remainder, we investigate the logical basis states individually, as tomographic reconstruction cannot be done fault-tolerantly.

We now address the critical question of how well our implementation of the repetition code protects logical states over many cycles. The process flow is illustrated in Fig. 4a. We start by initializing the data qubits in either of the logical basis states: $|0_L\rangle = |0..0\rangle$ or $|1_L\rangle = |1..1\rangle$. We then run the repetition code algorithm for k cycles, and finish by measuring the state of all data qubits. We repeat this 90,000 times to gather statistics. The classical measurement results are converted into detection events, which are processed using minimum-weight perfect matching to generate corrections (see Supplementary Information). These corrections are then applied to the measured data qubit output

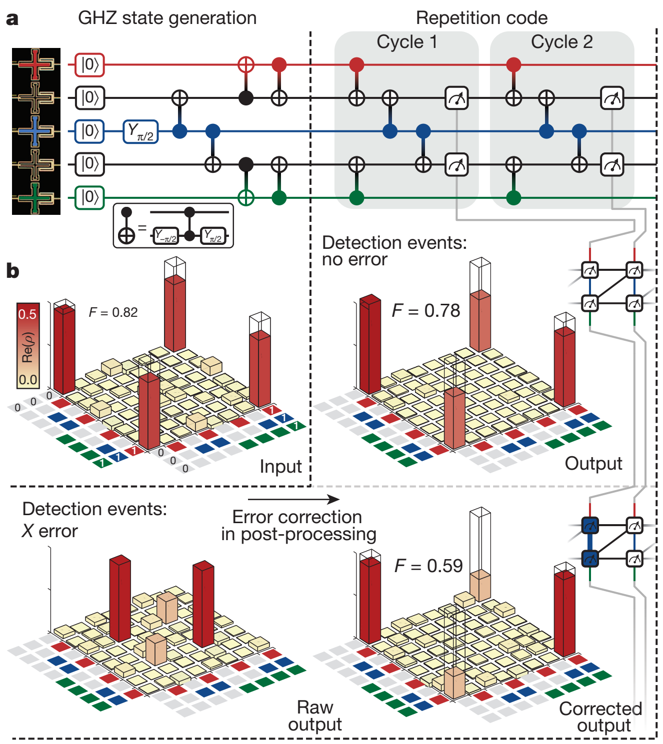


Figure 3 | Protecting the GHZ state from bit-flip errors. **a**, Quantum circuit for generating the GHZ state and two cycles of the repetition code. CNOT gates are physically implemented with controlled-phase (CZ) and single qubit gates. **b**, Quantum state tomography on the input (top left 'Input', left of black dashed line), and after the repetition code conditional on the detection events (between black dashed lines): we input a GHZ state with a fidelity (F) of 82%, and find, for the case of no detection events (top right 'Output', above grey dashed line), a 78% fidelity GHZ state. For the detection event connecting both measurement qubits (bottom left 'Raw output', below grey dashed line), indicating a likely bit-flip error on the central data qubit, we find that through correcting in post-processing by exchanging matrix elements we recover the major elements of the diagonal (bottom right 'Corrected output'). We also recover non-zero off-diagonal elements, indicating some bit-flip errors are coherent. Real parts are shown, $|\mathrm{Im}(\rho)| < 0.03$.

to see if the input is recovered. Owing to the topological nature of errors in space and time, we either recover the logical state, or the bit-wise inverse (see Supplementary Information). The fidelity of the repetition code algorithm is defined by the success rate of this recovery. In our system, qubits naturally relax to $|0\rangle$, intrinsically making $|0_L\rangle$ more robust than $|1_L\rangle$. To balance these errors and to increase the worst-case lifetime of the system, we apply physical bit-flips to each data qubit at the end of each cycle. This logical flip is compensated for in software. In principle, bit-flip gates can be applied to a single physical qubit to reduce susceptibility to energy relaxation by at best a factor of two. However, this will introduce coherent gate errors. In order to quantify the reduction of logical errors with system size n, we have implemented the repetition code with five and nine qubits in total, corresponding to first- and second-order fault-tolerance.

In Fig. 4b we show the fidelity of the repetition code as a function of the number of cycles for five (blue) and nine (red) qubits. We also plot the probability of a $|1\rangle$ state idling for the same duration, averaged over the five data qubits (black). This allows for a direct comparison of single physical qubit error with the multi-qubit logical error. We find a reduced error of logical states after eight cycles as compared to a physical qubit; by a factor of 2.7 for five qubits and 8.5 for nine qubits. We also see a non-exponential fidelity decay for logical states, due to an increasing error rate with cycle number (Fig. 4c). This effect can be explained primarily by energy relaxation of measurement qubits, with a small contribution

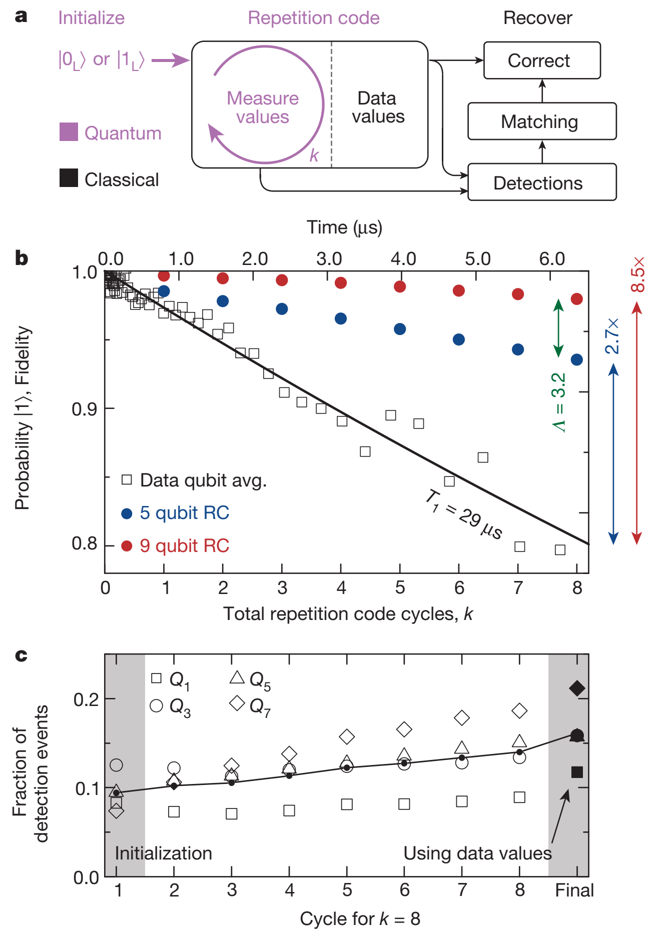


Figure 4 | Logical state preservation with the repetition code. **a**, Information flowchart of the repetition code. The data qubits are initialized into $|0_L\rangle$ or $|1_L\rangle$, and the repetition code is repeated k times. In post-processing, the measurement qubit outcomes are converted into detection events and matched to find likely errors, see Fig. 2. A successful recovery converts the measured data qubit state into the input state. **b**, Memory fidelity versus time and cycles for a single physical qubit (black) and the five- (blue) and nine- (red) qubit repetition code. Note that energy relaxation decays from a fidelity of 1 to 0, whereas the repetition code decays from a fidelity of 1 to 0.5. Five qubit data sampled from nine qubit data, see Supplementary Information. The average physical qubit lifetime ('data qubit avg.') is $T_1 = 29\ \mu s$, and after eight cycles we see an improvement in error rate by a factor of 2.7 (blue arrow at right) for five qubits ('5 qubit RC'), and 8.5 (red arrow at right) for nine qubits ('9 qubit RC') when using the repetition code. This indicates a Λ parameter of 3.2 (green arrow) for our system after eight cycles. **c**, Average number of detection events per measurement qubit (open symbols), versus cycle number, for experiments consisting of eight cycles. We see an increasing average rate of detection events (black line) with increasing cycle number. This can be attributed to the statistically increasing number of odd parity $\hat{Z}\hat{Z}$ measurements, see text. Grey regions indicate initialization and final data qubit measurement.

from state leakage. Initial logical states of all 0s or 1s have even parity for all $\hat{Z}\hat{Z}$ operators, maintaining the initial measurement qubit $|0\rangle$ state. A bit-flip error on a data qubit, statistically more likely with increasing cycle number, will cause the nearby $\hat{Z}\hat{Z}$ operators to have odd parity. This will flip measurement qubits between the $|0\rangle$ and $|1\rangle$ state at each cycle, making them susceptible to energy relaxation and hence increasing the rate of detection events (see Supplementary Information).

Figure 4 demonstrates state preservation through error correction. We emphasize that we correct errors that intrinsically arise from the environment. Additionally, we see larger repetition codes leading to greater error suppression. This is evidence for the system operating with fidelities above the repetition code threshold. As the error rates

grow with cycle number, the many-cycle behaviour of the repetition code must be explored to ensure that the system remains above threshold. The ratio of the errors for the $n = 1$ and $n = 2$ cases after eight cycles suggests $\Lambda = 3.2$, but larger system sizes are needed to infer this accurately for large n and verify that the logical error rate is suppressed exponentially as $\varepsilon_{\mathrm{logical}} \propto 1/\Lambda^{n+1}$, as desired.

Our demonstration that information can be stored with lower error in logical states than in single physical qubits shows that the basic physical processes required to implement surface code error correction are technologically feasible. We hope that our work will help to accelerate research into the many outstanding challenges that remain, such as the development of two-dimensional qubit arrays with scalable wiring and four-qubit QND parity checks, improving gate and measurement fidelities[27], and investigating the many-cycle behaviour of error correction schemes.

Received 22 November 2014; accepted 27 January 2015.

1. Shor, P. W. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52,** R2493–R2496 (1995).
2. Calderbank, A. R. & Shor, P. W. Good quantum error-correcting codes exist. *Phys. Rev. A* **54,** 1098–1105 (1996).
3. Steane, A. M. Error correcting codes in quantum theory. *Phys. Rev. Lett.* **77,** 793–797 (1996).
4. Bravyi, S. B. & Kitaev, A. Y. Quantum codes on a lattice with boundary. Preprint at http://arxiv.org/abs/quantph/9811052 (1998).
5. Raussendorf, R. & Harrington, J. Fault-tolerant quantum computation with high threshold in two dimensions. *Phys. Rev. Lett.* **98,** 190504 (2007).
6. Raussendorf, R., Harrington, J. & Goyal, K. Topological fault-tolerance in cluster state quantum computation. *New J. Phys.* **9,** 199 (2007).
7. Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A* **86,** 032324 (2012).
8. Cory, D. G. *et al.* Experimental quantum error correction. *Phys. Rev. Lett.* **81,** 2152–2155 (1998).
9. Knill, E., Laflamme, R., Martinez, R. & Negrevergne, C. Benchmarking quantum computers: the five qubit error correcting code. *Phys. Rev. Lett.* **86,** 5811–5814 (2001).
10. Chiaverini, J. *et al.* Realization of quantum error correction. *Nature* **432,** 602–605 (2004).
11. Reed, M. *et al.* Realization of three-qubit quantum error correction with superconducting circuits. *Nature* **482,** 382–385 (2012).
12. Chow, J. M. *et al.* Implementing a strand of a scalable fault-tolerant quantum computing fabric. *Nature Commun.* **5,** 4015 (2014).
13. Riste, D. *et al.* Detecting bit-flip errors in a logical qubit using stabilizer measurements. Preprint at http://arxiv.org/abs/1411.5542 (2014).
14. Nigg, D. *et al.* Quantum computations on a topologically encoded qubit. *Science* **345,** 302–305 (2014).
15. Córcoles, A. *et al.* Detecting arbitrary quantum errors via stabilizer measurements on a sublattice of the surface code. Preprint at http://arxiv.org/abs/1410.6419 (2014).
16. Schindler, P. *et al.* Experimental repetitive quantum error correction. *Science* **332,** 1059–1061 (2011).
17. Sun, L. *et al.* Tracking photon jumps with repeated quantum non-demolition parity measurements. *Nature* **511,** 444–448 (2014).
18. Bombin, H. & Martin-Delgado, M. Quantum measurements and gates by code deformation. *J. Phys. A* **42,** 095302 (2009).
19. Barends, R. *et al.* Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* **508,** 500–503 (2014).
20. Jeffrey, E. *et al.* Fast accurate state measurement with superconducting qubits. *Phys. Rev. Lett.* **112,** 190504 (2014).
21. Barends, R. *et al.* Coherent Josephson qubit suitable for scalable quantum integrated circuits. *Phys. Rev. Lett.* **111,** 080502 (2013).
22. Fowler, A. G., Sank, D., Kelly, J., Barends, R. & Martinis, J. M. Scalable extraction of error models from the output of error detection circuits. Preprint at http://arxiv.org/abs/1405.1454 (2014).
23. Edmonds, J. Paths, trees, and flowers. *Can. J. Math.* **17,** 449–467 (1965).
24. Edmonds, J. Maximum matching and a polyhedron with 0,1-vertices. *J. Res. Natl Bur. Stand. B* **69,** 125–130 (1965).
25. Fowler, A. G. Minimum weight perfect matching of fault-tolerant topological quantum error correction in average O(1) parallel time. *Quant. Inform. Comput.* **15,** 0145–0158 (2015).
26. Gühne, O. & Seevinck, M. Separability criteria for genuine multiparticle entanglement. *New J. Phys.* **12,** 053002 (2010).
27. Harty, T. *et al.* High-fidelity preparation, gates, memory and readout of a trapped-ion quantum bit. *Phys. Rev. Lett.* **113,** 220501 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. N. Korotkov and D. L. Moehring for discussions, and P. Duda for help with photomasks and photolithography. This work was supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through Army Research Office grants W911NF-09-1-0375 and W911NF-10-1-0334. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI or the US Government. Devices were made at the UC Santa Barbara Nanofabrication Facility, a part of the US NSF-funded National Nanotechnology Infrastructure Network, and at the NanoStructures Cleanroom Facility.

Author Contributions J.K. and R.B. designed the sample and performed the experiment. A.G.F. and J.M.M. designed the experiment. J.K., R.B. and A.M. fabricated the sample. A.G.F., J.K. and R.B. analysed the data. J.K., R.B., A.G.F. and J.M.M. co-wrote the manuscript. All authors contributed to the fabrication process, experimental set-up and manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.K. (julian@physics.ucsb.edu) or J.M.M. (martinis@physics.ucsb.edu).

doi:10.1038/nature14253

Liquid-based gating mechanism with tunable multiphase selectivity and antifouling behaviour

Xu Hou[1,2], Yuhang Hu[1], Alison Grinthal[1], Mughees Khan[2] & Joanna Aizenberg[1,2]

Living organisms make extensive use of micro- and nanometre-sized pores as gatekeepers for controlling the movement of fluids, vapours and solids between complex environments. The ability of such pores to coordinate multiphase transport, in a highly selective and subtly triggered fashion and without clogging, has inspired interest in synthetic gated pores for applications ranging from fluid processing to 3D printing and lab-on-chip systems[1–10]. But although specific gating and transport behaviours have been realized by precisely tailoring pore surface chemistries and pore geometries[6,11–17], a single system capable of controlling complex, selective multiphase transport has remained a distant prospect, and fouling is nearly inevitable[11,12]. Here we introduce a gating mechanism that uses a capillary-stabilized liquid as a reversible, reconfigurable gate that fills and seals pores in the closed state, and creates a non-fouling, liquid-lined pore in the open state. Theoretical modelling and experiments demonstrate that for each transport substance, the gating threshold—the pressure needed to open the pores—can be rationally tuned over a wide pressure range. This enables us to realize in one system differential response profiles for a variety of liquids and gases, even letting liquids flow through the pore while preventing gas from escaping. These capabilities allow us to dynamically modulate gas–liquid sorting in a microfluidic flow and to separate a three-phase air–water–oil mixture, with the liquid lining ensuring sustained antifouling behaviour. Because the liquid gating strategy enables efficient long-term operation and can be applied to a variety of pore structures and membrane materials, and to micro- as well as macroscale fluid systems, we expect it to prove useful in a wide range of applications.

Our hypothesis that a liquid-filled pore could provide a unified gating strategy derives from the idea that a liquid stabilized inside a micropore offers a unique combination of dynamic and interfacial behaviours, and is inspired by nature's use of fluids as reconfigurable gates. Microscale stomata and xylem control air, water and microbe exchange in plants by using liquid to mechanically reconfigure the pore[18]. The nuclear pore is directly lined with disordered, fluid-like proteins that are believed not only to regulate differential transport of a wide range of cargos, but also to completely prevent fouling[19]. Most interestingly, micropores between air sacs in the lung are filled with liquid that has been proposed to reversibly reconfigure inside the pore to create an open, liquid-lined pathway in response to pressure gradients[20].

Figure 1 contrasts the gating mechanisms in a conventional pore and in a liquid-filled pore. In the case of conventional nano- or micropores (Fig. 1a), gases will flow through passively regardless of pore shape and surface chemistry, whereas liquids will enter the pore once the applied pressure reaches a critical value dictated by the balance of surface interactions, pore geometry and surface tension[21–23]. If the pores are filled with a strongly wetting liquid that completely seals them and forms a contiguous coating along the adjacent surface (Fig. 1b), gases and liquids must deform the interface of the pore-filling liquid to enter the pore and will require different pressure thresholds to do so. As long as the pore-filling liquid's affinity for the solid is stronger than that of the transport substance, the pore liquid will part to form an open, liquid-lined pathway while remaining adherent to the pore walls and adjacent surfaces, such that the transport substance, as long as it is immiscible, will flow through without contacting any solid surfaces. Unlike with bare pores, the transport substance will thus contact only the liquid in either the open or the closed state, preventing fouling both inside[24] and around[25] the pore. Because this dynamic gating mechanism involves structural reconfiguration, rather than expulsion, of the pore liquid, the pore will stay open only as long as the transport substance is flowing and will be thermodynamically primed to close as soon as the pressure drops below the threshold.

This rationale can be used to design gated transport systems having a wide variety of pore sizes, pore geometries, surface chemistries and gating liquids. We illustrate this by using membrane materials with different pore structures (Fig. 2a and Supplementary Fig. 1) and chemistries, ranging from hydrophobic polytetrafluoroethylene (PTFE), poly(vinylidene fluoride) and polypropylene (Supplementary Table 1) to hydrophilic nylon. Air flows through all of these materials in their unmodified state at zero pressure, but adding a gating liquid creates a substantial critical gating pressure that has a distinct value for each membrane (Fig. 2b). Because the liquid-gated pore can open and close in response to different conditions and transport substances, gas and liquid flows can be differentially controlled in a single system. For example, filling PTFE pores with a low-surface-energy liquid simultaneously generates

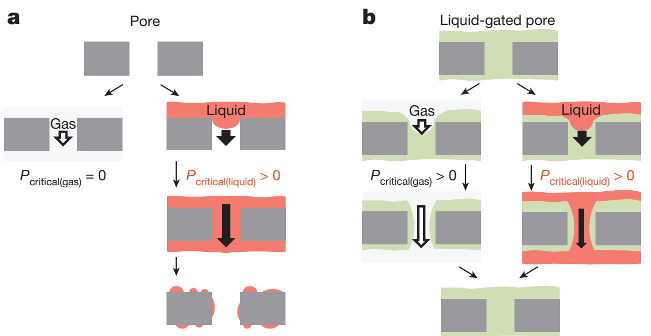


Figure 1 | Hypothesis for gating a pore by liquid reconfiguration. **a**, For a solid nano- or micropore (significantly larger than the molecular scale), transport of gases (light grey) is uncontrolled and occurs even at zero pressure, whereas transport of liquids (red) depends on the meniscus formation defined by the interactions with the solid surface and therefore occurs at specific finite pressure. The system is prone to fouling. **b**, If the pore is filled with a stably held liquid (green), flow of both gases and liquids will be gated by pressure-induced deformation of the gating liquid surface. In the open state, the gating liquid will reversibly reconfigure to form a liquid-lined pore. Each transport substance will have a specific critical pressure based on its ability to overcome the capillary pressure at the liquid-gas or liquid-liquid interface, and the liquid-lined pore will prevent contact with the solid. When the pressure is released, a non-fouled pore returns to its original liquid-filled state. The liquid-based gating mechanism provides a unified strategy for selective, responsive, tunable and antifouling multiphase transport.

[1]School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. [2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts 02138, USA.

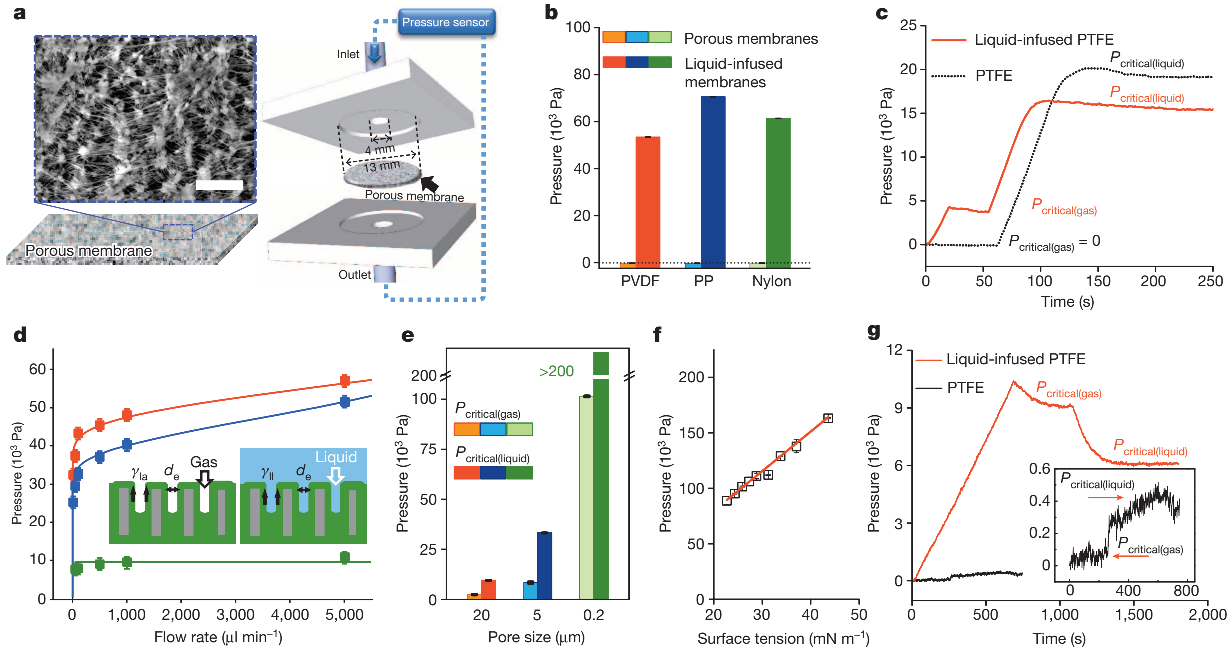


Figure 2 | Design and rational tuning of gating systems with differentially controlled gas and liquid transport. **a**, Left: scanning electron micrograph of a typical porous membrane. Scale bar, 5 μm. Right: sketch of the pressure measurement set-up. **b**, Critical pressure required for gas to flow through porous materials of various chemistries and pore structures (average pore size, 0.45 μm), with and without a gating liquid. For each material, $P_{critical(gas)} = 0$ without the gating liquid and $P_{critical(gas)} > 0$ with the gating liquid. PP, polypropylene; PVDF, poly(vinylidene fluoride). **c**, Liquid-based gating creates different critical pressures for gas and water within a single system. Importantly, the critical pressure for water transport through a liquid-filled membrane is less than that for transport through the same membrane without liquid. **d**, The predictive model (lines) agrees with experimentally obtained critical pressures (squares) over a range of flow rates for air through liquid-gated pores (green), water through liquid-gated pores (blue), and water through pores without a liquid gate (red). Inset, schematic model for determining critical pressure for gases and liquids. In both cases, the gating pressure is a function of the pore geometry and interfacial (gas–liquid or liquid–liquid) tension. **e**, Differential tuning of absolute and relative critical pressures for air and water to flow through liquid-gated pores with different pore sizes. **f**, Systematic tuning of the gating threshold for air transport by infusing porous nylon membranes with a series of gating liquids of different surface tensions, achieved by mixing water and ethanol in different ratios. **g**, Critical pressures for air (black) and ethanol (red) through liquid-gated pores. The difference in interfacial tensions leads to an unusual situation in which the critical pressure for liquid is lower than that for gas. Inset: without the gating liquid, the critical pressures for both air and ethanol are negligible. All error bars, 1 s.d.

a gating pressure for air and alters—in this case lowers—the critical pressure for water, such that both substances can be transported through the membrane in succession according to their distinct thresholds (Fig. 2c and Supplementary Fig. 2).

Gating for gases and liquids is expected to occur via the same capillary mechanism, where the critical pressure will be the pressure needed to deform the surface of the pore-filling liquid (Fig. 2d, insets). For a gas, the pressure that must be overcome is the Laplace pressure, $4\gamma_{la}/d_e$, where γ_{la} is the surface tension of the pore-filling liquid and d_e is the average effective pore size[26]. These parameters are sufficient to predict the experimentally observed critical pressure for air in a 5 μm porous membrane infused with a low-surface-energy liquid (Fig. 2d, green). For a liquid, the gating pressure will depend on d_e and on the liquid–liquid interfacial tension γ_{ll}; in fact, this relationship[27] is used to characterize membrane porosity by observing the irreversible expulsion of the pore-filling liquid[26,28]. But in our case, where we aim for the gating liquid to reversibly reconfigure in place so that the transport liquid continuously flows through a liquid-lined pore, the working pressure will also depend on the flow rate Q and viscosity μ of the transport fluid[29], as $\Delta P \propto Q\mu/k$. In this expression, k is the permeability of the membrane, which is related to the pore structure and size and also depends on the transmembrane pressure or flow rate[29,30]:

$$k = \frac{\Phi}{32\tau^2}\int_{2\gamma_{ll}/\Delta P}^{\infty} \frac{X^2}{\sigma\sqrt{2\pi}}\, e^{-(X-d)^2/2\sigma^2}\, dX$$

Here Φ is the porosity, τ the tortuosity, d the mean pore size and σ the standard deviation of a porous membrane with distributed pore sizes (Supplementary Information). This relationship accurately predicts the gating pressure for water at a series of flow rates, both with and without the pore-filling liquid (Fig. 2d, blue and red), and allows us to quantitatively determine how the performance of the system depends on the pore size, geometry and gating liquid properties (Supplementary Fig. 3 and Supplementary Tables 2 and 3).

We can use a variety of membranes with different pore sizes to tune the absolute and relative critical pressures for gases and liquids over at least two orders of magnitude, from less than 10 kPa to more than 200 kPa (Fig. 2e and Supplementary Fig. 4). For a given membrane material and pore size, the gating pressure is finely tuned by using pore-filling liquids with systematically varying surface tensions (Fig. 2f, Supplementary Fig. 5 and Supplementary Table 4). We can even adjust the system so that the critical pressure for air exceeds that for the liquid, in this case ethanol (Fig. 2g). In all cases, the observed gating thresholds are independent of whether gas or liquid is flowed first (Supplementary Fig. 6), and are stable over time and after cyclic alteration of gas and liquid (Supplementary Fig. 7).

This combination of differential tunability and reversible opening and closing enables fast and repeatable control over multiphase flows in both microfluidic and macrofluidic systems by simply adjusting the system pressure. For example, we can produce distinct air–water streams by incorporating a liquid-gated porous membrane into a port in a microfluidic channel and setting the critical pressures for air and water to 2.3 and 9.8 kPa, respectively (Fig. 3a, Supplementary Figs 8–11 and Supplementary Video 1). Below 2.3 kPa, neither substance flows through the port (Fig. 3a, first panel). Between the two critical pressures, only air flows through the port, and gas-free water continues past it (Fig. 3a,

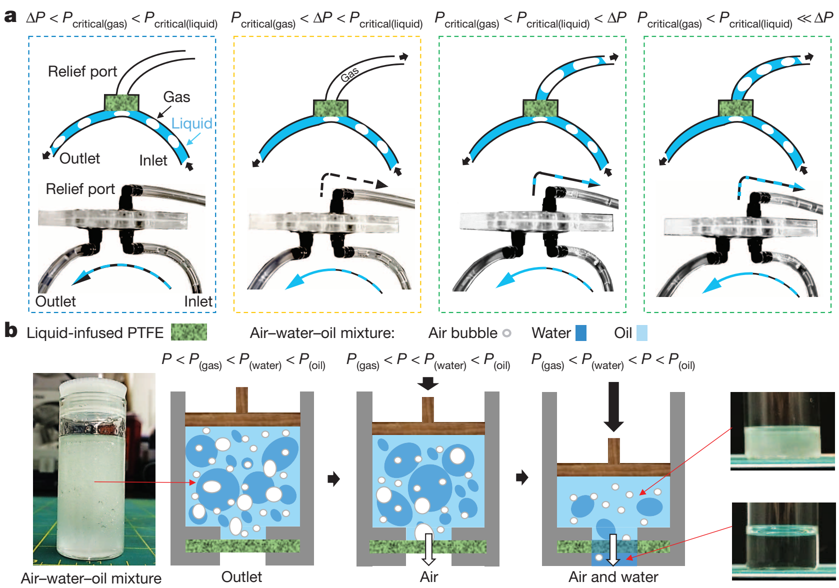


Figure 3 | Sorting of multiphase mixtures by liquid-gated pores. **a**, Rapid pressure-tunable sorting of multiphase flows in a microfluidic system. Top: sketches; bottom: snapshots of experiment. A liquid-infused porous material incorporated into a port along a microfluidic channel enables a series of distinct pressure-dependent scenarios for a mixed gas–liquid flow. In the sketches, the membrane is shown in green and the transported liquid is shown in blue; the blue arrows depict the transport of liquid, dashed black and blue arrows correspond to the transport of a mixed gas–liquid phase, and dashed black and white arrows depict the transport of gas only. First panel: at pressures below both $P_{\mathrm{critical(gas)}}$ and $P_{\mathrm{critical(liquid)}}$, nothing flows through the port. Second panel: above $P_{\mathrm{critical(gas)}}$ and below $P_{\mathrm{critical(liquid)}}$, only the gas flows through the port, and degassed liquid continues through the channel beyond the port. Third and fourth panels: at pressures above both $P_{\mathrm{critical(gas)}}$ and $P_{\mathrm{critical(liquid)}}$, both phases cross the port and only liquid continues through the channel beyond the port. Note that above both critical pressures, the liquid/gas ratio of the mixture that crosses the port increases with increasing pressure. Data shown are for an alternating air–liquid flow, with $P_{\mathrm{critical(gas)}} = 2.3$ kPa and $P_{\mathrm{critical(liquid)}} = 9.8$ kPa. **b**, A three-phase mixture of air, water and crude oil is progressively separated by the liquid-infused porous membrane.

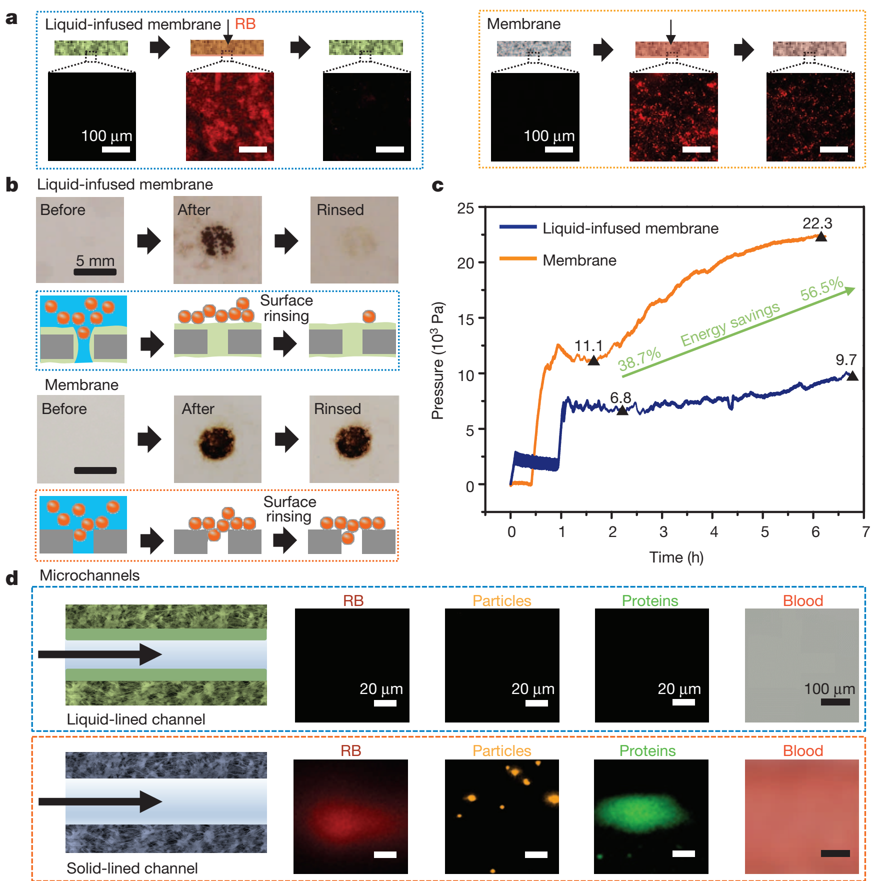


Figure 4 | Antifouling transport and separation of complex substances. **a**, Real-time confocal images of a liquid-infused porous material (left) or of the porous material without the gating liquid (right) before, during and after flowing an aqueous solution of rhodamine B dye (RB). Detention of dye after flow is observable only in the PTFE membrane without the gating liquid. **b**, Flowing a suspension of 4-benzoylamino-2,5-diethoxybenzenediazonium chloride hemi(zinc chloride) salt particles (1,000 μl min^{-1}) through a liquid-infused porous material (top) leaves salt particles suspended on the liquid surface after the pore closes (top, centre), and the particles are easily collected by a gentle surface rinse (top, right). The bare membrane (bottom) traps the salts both in and around the pores (bottom, centre) and resists repeated rinsing (bottom, right). **c**, The pressure required to transport a suspension of 4-benzoylamino-2,5-diethoxybenzenediazonium chloride hemi(zinc chloride) salt particles in water through a liquid-infused PTFE membrane is initially 38.7% lower than the pressure required for transport through the bare membrane. This percentage difference increases to about 56.5% after 4.5 h operation at a flow rate of 50 μl min^{-1} owing to fouling of the bare membrane. The energy saving rate is $(P_{\mathrm{PTFE}} - P_{\mathrm{PTFE\text{-}Krytox\ 103}})/P_{\mathrm{PTFE}}$, where Krytox 103 is the pore-filling liquid. **d**, A liquid-lined PTFE microchannel resists fouling (top) by an RB solution, suspended microparticles and a protein solution (fluorescent images) and by blood (optical image), whereas fouling residue is observed inside the bare PTFE channel for each substance (bottom).

second panel). Above 9.8 kPa, both air and water flow through the port (Fig. 3a, third panel), with their precise balance responding to graded pressure changes (Fig. 3a, fourth panel) because pressure increases the flow rate through the port significantly for liquids but only negligibly for gases (Supplementary Information and Supplementary Fig. 12). The robustness of the gating behaviour indicates that the pores can sustain many rapid opening–closing cycles while remaining faithful to the original gating pressures, allowing the system to be operated continuously for at least several days (Supplementary Video 1). We also explore more complex fluid handling using these principles: by setting a distinct threshold for each component of a three-phase gas–liquid–liquid mixture, we use pressure to actively adjust and control selective fluid flow through the liquid-gated membrane to collect the different phases while preventing the escape of any component (Fig. 3b and Supplementary Fig. 13).

The liquid-gating strategy further suppresses fouling, as illustrated by the real-time confocal images in Fig. 4a, which show that a rhodamine B dye solution leaves no trace on liquid-gated membranes once the flow stops, whereas bare pores retain substantial dye (Supplementary Figs 14–15). Lining the membrane pores and outer surfaces with liquid thus indeed prevents the transport substance from contacting the solid (Fig. 1), enabling the separation of suspensions containing particles larger than the pore diameter without risk of fouling. This is further illustrated in Fig. 4b, where salt particles are found only on top of the closed pores following flow and are collected by a gentle surface rinse; the bare membrane traps the particles in and around the pores and resists repeated rinsing (Supplementary Fig. 16). The integration of antifouling behaviour and pressure-controlled flow properties in one system not only provides the flexibility to have a lower working pressure for liquid transport than with conventional membranes, but also avoids the common problem of working pressure build-up caused by pore clogging and uncontrolled fouling. The combination of these effects can lead to significant energy savings in long-term operation (Fig. 4c; see Supplementary Information for discussion).

A liquid lining prevents fouling not only inside nano- and micropores, but also in much longer microfluidic channels. As illustrated in Fig. 4d, a microfluidic channel lined with a liquid-filled porous membrane shows no retention of rhodamine B, fluorescent microparticles, fluorescent protein or blood, whereas state-of-the-art but conventional non-fouling microfluidic channels retain substantial residue. By using liquid-lined channels and pores, it should therefore be possible to design microfluidic systems that resist fouling and enable tunable gated flow of complex multiphase substances.

The dynamic reconfiguration of a liquid lining as a means to reversibly open and close membrane pores provides an attractive and tunable gating mechanism that integrates chemical and physical selectivity for gas- and liquid-phase transport substances with non-fouling behaviour and energy-efficient operation. We anticipate that these capabilities, combined with the longevity of the membrane systems, the range of materials and geometries that can be used in their construction, and their applicability to both macro- and microfluidic operation, will be of benefit in fields ranging from water treatment and biomedical fluid processing to 3D printing, microscale reactor operation and beyond.

Received 22 June 2014; accepted 15 January 2015.

1. Chen, P. C. & Xu, Z. K. Mineral-coated polymer membranes with superhydrophilicity and underwater superoleophobicity for effective oil/water separation. *Sci. Rep.* **3,** 2776 (2013).
2. Holt, J. K. *et al.* in *Proc. 4th IEEE Conf. Nanotechnol.* 110–112 (IEEE, 2004).
3. Peng, X. S., Jin, J., Nakamura, Y., Ohno, T. & Ichinose, I. Ultrafast permeation of water through protein-based membranes. *Nature Nanotechnol.* **4,** 353–357 (2009).
4. Paven, M. *et al.* Super liquid-repellent gas membranes for carbon dioxide capture and heart-lung machines. *Nature Commun.* **4,** 2512 (2013).
5. Faulkner-Jones, A. *et al.* Development of a valve-based cell printer for the formation of human embryonic stem cell spheroid aggregates. *Biofabrication* **5,** 015013 (2013).
6. Oh, K. W. & Ahn, C. H. A review of microvalves. *J. Micromech. Microeng.* **16,** R13–R39 (2006).
7. Kargov, A. *et al.* in *Proc. IEEE 9th Int. Conf. Rehabilitation Robotics* 182–186 (IEEE, 2005).
8. Edwards, D. A. *et al.* Large porous particles for pulmonary drug delivery. *Science* **276,** 1868–1872 (1997).
9. Kralj, J. G., Sahoo, H. R. & Jensen, K. F. Integrated continuous microfluidic liquid-liquid extraction. *Lab Chip* **7,** 256–263 (2007).
10. Karlsson, J. M. *et al.* Active liquid degassing in microfluidic systems. *Lab Chip* **13,** 4366–4373 (2013).
11. Ulbricht, M. Advanced functional polymer membranes. *Polymer* **47,** 2217–2262 (2006).
12. Lin, N. H., Kim, M. M., Lewis, G. T. & Cohen, Y. Polymer surface nano-structuring of reverse osmosis membranes for fouling resistance and improved flux performance. *J. Mater. Chem.* **20,** 4642–4652 (2010).
13. Powell, M. R., Cleary, L., Davenport, M., Shea, K. J. & Siwy, Z. S. Electric-field-induced wetting and dewetting in single hydrophobic nanopores. *Nature Nanotechnol.* **6,** 798–802 (2011).
14. Yameen, B. *et al.* Ionic transport through single solid-state nanopores controlled with thermally nanoactuated macromolecular gates. *Small* **5,** 1287–1291 (2009).
15. Wen, Y. Q. *et al.* DNA-based intelligent logic controlled release systems. *Chem. Commun.* **48,** 8410–8412 (2012).
16. Adrus, N. & Ulbricht, M. Novel hydrogel pore-filled composite membranes with tunable and temperature-responsive size-selectivity. *J. Mater. Chem.* **22,** 3088–3098 (2012).
17. Nair, R. R., Wu, H. A., Jayaram, P. N., Grigorieva, I. V. & Geim, A. K. Unimpeded permeation of water through helium-leak–tight graphene-based membranes. *Science* **335,** 442–444 (2012).
18. Stroock, A. D., Pagay, V. V., Zwieniecki, M. A. & Holbrook, N. M. The physicochemical hydrodynamics of vascular plants. *Annu. Rev. Fluid Mech.* **46,** 615–642 (2014).
19. Peleg, O. & Lim, R. Y. H. Converging on the function of intrinsically disordered nucleoporins in the nuclear pore complex. *Biol. Chem.* **391,** 719–730 (2010).
20. Namati, E., Thiesse, J., de Ryk, J. & McLennan, G. Alveolar dynamics during respiration: are the pores of Kohn a pathway to recruitment? *Am. J. Respir. Cell Mol. Biol.* **38,** 572–578 (2008).
21. Winther-Jensen, B., Winther-Jensen, O., Forsyth, M. & Macfarlane, D. R. High rates of oxygen reduction over a vapor phase-polymerized PEDOT electrode. *Science* **321,** 671–674 (2008).
22. Mohns, J. & Kunnecke, W. Flow-analysis with membrane separation and time-based sampling for ethanol determination in beer and wine. *Anal. Chim. Acta* **305,** 241–247 (1995).
23. Liu, C. C., Thompson, J. A. & Bau, H. H. A membrane-based, high-efficiency, microfluidic debubbler. *Lab Chip* **11,** 1688–1693 (2011).
24. Yusko, E. C. *et al.* Controlling protein translocation through nanopores with bio-inspired fluid walls. *Nature Nanotechnol.* **6,** 253–260 (2011).
25. Wong, T. S. *et al.* Bioinspired self-repairing slippery surfaces with pressure-stable omniphobicity. *Nature* **477,** 443–447 (2011).
26. Mietton-Peuchot, M., Condat, C. & Courtois, T. Use of gas-liquid porometry measurements for selection of microfiltration membranes. *J. Membr. Sci.* **133,** 73–82 (1997).
27. Zhang, C. Y., Oostrom, M., Wietsma, T. W., Grate, J. W. & Warner, M. G. Influence of viscous and capillary forces on immiscible fluid displacement: pore-scale experimental study in a water-wet micromodel demonstrating viscous and capillary fingering. *Energy Fuels* **25,** 3493–3505 (2011).
28. Germic, L. *et al.* Characterization of polyacrylonitrile ultrafiltration membranes. *J. Membr. Sci.* **132,** 131–145 (1997).
29. Biot, M. A. General theory of three-dimensional consolidation. *J. Appl. Phys.* **12,** 155–164 (1941).
30. Purcell, W. R. Capillary pressures - their measurement using mercury and the calculation of permeability therefrom. *J. Petrol. Technol.* **1,** 39–48 (1949).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported in part by the Advanced Research Projects Agency-Energy (ARPA-E), US Department of Energy, under award number DE-AR0000326. We thank M. Aizenberg, R. T. Blough and X. Y. Chen for discussions; A. B. Tesler for assistance with the scanning electron microscopy; and T. S. Wong, B. D. Hatton and R. A. Belisle for assistance with antifouling experiments.

Author Contributions X.H. and J.A. designed the liquid-infused porous materials and the experiments. X.H. and M.K. carried out the experiments. All authors analysed data. Y.H. built the mathematical model. All authors interpreted data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A. (jaiz@seas.harvard.edu).

doi:10.1038/nature14255

The direct arylation of allylic sp^3 C–H bonds via organic and photoredox catalysis

James D. Cuthbertson[1] & David W. C. MacMillan[1]

The direct functionalization of unactivated sp^3 C–H bonds is still one of the most challenging problems facing synthetic organic chemists. The appeal of such transformations derives from their capacity to facilitate the construction of complex organic molecules via the coupling of simple and otherwise inert building blocks, without introducing extraneous functional groups. Despite notable recent efforts[1], the establishment of general and mild strategies for the engagement of sp^3 C–H bonds in C–C bond forming reactions has proved difficult. Within this context, the discovery of chemical transformations that are able to directly functionalize allylic methyl, methylene and methine carbons in a catalytic manner is a priority. Although protocols for direct oxidation and amination of allylic C–H bonds (that is, C–H bonds where an adjacent carbon is involved in a C=C bond) have become widely established[2,3], the engagement of allylic substrates in C–C bond forming reactions has thus far required the use of pre-functionalized coupling partners[4]. In particular, the direct arylation of non-functionalized allylic systems would enable access to a series of known pharmacophores (molecular features responsible for a drug's action), though a general solution to this long-standing challenge remains elusive. Here we report the use of both photoredox and organic catalysis to accomplish a mild, broadly effective direct allylic C–H arylation. This C–C bond forming reaction readily accommodates a broad range of alkene and electron-deficient arene reactants, and has been used in the direct arylation of benzylic C–H bonds.

While the well-known Heck reaction can be employed to generate allyl-substituted aromatics with high levels of efficiency via π-addition, olefin transposition sequences, there are few known examples of direct allylic arylation via the functionalization of C–H bonds[5,6]. Indeed, a literature survey reveals an isolated report on the Fe-catalysed coupling of Grignard reagents with simple olefins as the only transition-metal-mediated allylic arylation reported to date[7]. Given the state of the art of allylic C–H arylation (or for allylic C–C bond formation in general), we considered that a mild and widely applicable solution to this challenge would be useful.

The rapidly growing field of visible-light-mediated photoredox catalysis offers a valuable platform for the design or discovery of new synthetic transformations[8–10]. The ability of photoredox catalysts to act as both strong oxidants and reductants upon irradiation with low-energy visible light has enabled the invention of a series of useful bond constructions, previously thought to be unattainable via conventional pathways. In our laboratory, the synergistic merger of visible-light photoredox catalysis with organocatalysis has been instrumental in the development of a number of methods for the direct functionalization of unactivated sp^3 C–H bonds[11–13]. Using this powerful dual-catalysis paradigm, we recently reported the direct arylation of benzylic ethers via a C–H abstraction mechanism that proceeds through the heterocoupling of two catalytically generated radical species[14]. With this mechanistic blueprint in hand, we recently considered whether the combination of photoredox catalysis and organocatalysis could provide a solution to the long-standing and more significant challenge of direct allylic arylation (Fig. 1). Given that the allylic C–H bonds of simple alkenes are relatively weak (the allylic C–H bond dissociation energy (BDE) in cyclohexene is 83.2 kcal mol^{-1})[15], we hypothesized that olefinic substrates would undergo hydrogen atom abstraction using our photoredox conditions to generate transient allylic radicals, which thereafter would participate in a hetero radical–radical coupling pathway with arene radical anions generated *in situ*.

A detailed mechanism for the envisioned fragment coupling is depicted in Fig. 2. Upon irradiation with low-energy visible light (for example, from a 26 W household compact fluorescent lamp (CFL)), the iridium complex Ir(ppy)$_3$ (**1**) (ppy = 2-phenylpyridine) is known to undergo a metal-to-ligand charge transfer and intersystem crossing to generate the long-lived excited state Ir^{III} species **2** (lifetime τ = 1,900 ns)[16], which is a strong reductant: the reduction potential $E_{1/2}^{IV/*III}$ is −1.73 V versus the saturated calomel electrode (SCE) in MeCN, where *III indicates the excited state of the Ir^{III} catalyst[16]. It has long been established that the photoexcited state of this complex will readily undergo single-electron transfer with electron-deficient arenes such as 4-cyanopyridines and 1,4-dicyanobenzene (for example, reduction potential $E_{1/2}^{red}$ = −1.61 V versus SCE for 1,4-dicyanobenzene in MeCN)[17] to generate a persistent arene radical anion along with the oxidized photocatalyst **3**. Though the Ir^{IV} species **3** ($E_{1/2}^{IV/III}$ = +0.77 V versus SCE in MeCN)[16] is not likely to be sufficiently oxidizing to directly oxidize a typical thiol ($E_{1/2}^{red}$ = +1.12 V versus SCE for butanethiol in MeCN)[18], this event should be facilitated by the presence of a suitable base. The weakly acidic thiol (pK_a = 7.91 for methyl 2-mercaptoacetate in H$_2$O)[19] is deprotonated to yield the thiolate anion ($E_{1/2}^{red}$ = −0.85 V versus SCE for butanethiolate in MeCN)[18], which is readily oxidized by the photocatalyst. Based on the reported BDEs for typical thiols and allylic C–H bonds, we reasoned that the electrophilic thiyl radical **5** would readily abstract an allylic hydrogen atom from the alkene substrate (cyclohexene allylic C–H BDE = 83.2 kcal mol^{-1}

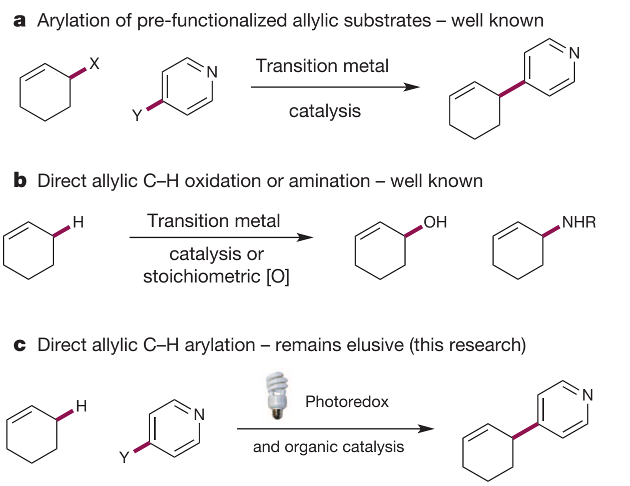


Figure 1 | The direct arylation of allylic C–H bonds via the synergistic merger of photoredox and organic catalysis. **a**, Arylation of allylic bonds is generally accomplished via transition-metal-catalysed couplings with pre-functionalized substrates. **b**, Installation of heteroatoms via direct allylic C–H functionalization is widely known. **c**, Direct C–H arylation is proposed via the synergistic merger of photoredox and organic catalysis.

[1]Merck Center for Catalysis at Princeton University, Princeton, New Jersey 08544, USA.

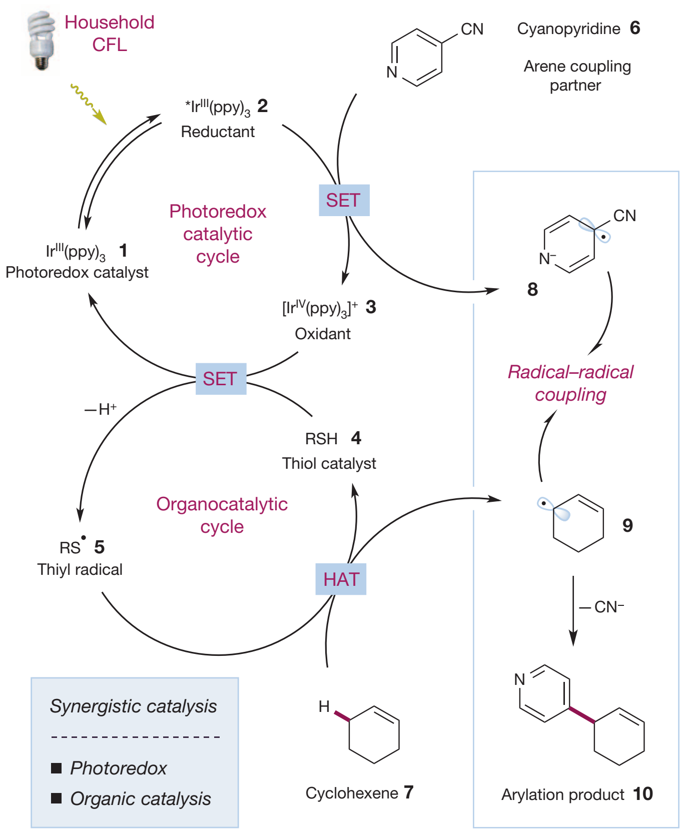


Figure 2 | Proposed mechanism for the direct arylation of allylic C–H bonds via photoredox and organic catalysis. The photoredox catalytic cycle is initiated via excitation of photocatalyst **1** to give the excited state **2**. Single-electron reduction of 4-cyanopyridine (**6**) via SET generates the radical anion **8** along with oxidant **3**. In the presence of a base, oxidant **3** is capable of initiating the organocatalytic cycle via oxidation of the thiol catalyst **4** to give the thiyl radical **5** along with the regenerated photocatalyst **1**. The thiyl radical **5** abstracts an allylic hydrogen atom (HAT, hydrogen atom transfer) from cyclohexene (**7**) to generate allylic radical **9**. A radical–radical coupling and subsequent elimination of cyanide serve to construct the new C–C bond and form the arylation product **10** while completing the two synergistic catalysis cycles.

versus S–H BDE = 87.0 kcal mol^{-1})[15,20] to provide allylic radical **9** along with regenerated organocatalyst **4**. At this time, an intermolecular radical–radical coupling would serve to forge the new C–C bond, with the resulting pyridienyl or cyclohexadienyl anion undergoing rapid rearomatization via elimination of cyanide.

We began our direct allylic C–H arylation studies with an evaluation of a range of photocatalysts, thiols, solvents, and bases in the presence of cyclohexene and 1,4-dicyanobenzene as representative coupling partners. We found that the use of light (from two 26-W CFLs) in the presence of Ir(ppy)$_3$, triisopropylsilanethiol, and K_2CO_3 enabled the desired aryl–allyl fragment coupling in excellent yield (87%) to furnish the desired arylcyclohexene. The necessity of each of the key reaction components—photocatalyst, thiol, and light—was demonstrated through a series of control experiments (Supplementary Information). While trace amounts of product were formed in the absence of thiol and photocatalyst[5], no reaction was observed upon exclusion of light. It should be noted that these optimal conditions employ standard household 26 W CFLs and proceed readily at room temperature with low loadings of both the photocatalyst (1 mol%) and organocatalyst (5 mol%).

With these optimized conditions in hand, we next examined the scope of olefins that can be employed in this direct allylic arylation reaction. As shown in Fig. 3a, this simple arylation protocol permits the direct coupling of electron-deficient arenes with a wide range of unfunctionalized alkenes. Importantly, both cyclic and acyclic olefins are readily accommodated in this transformation. For example, a series of simple cyclic alkenes coupled with 1,4-dicyanobenzene in excellent yields (**12**–**16**, 71–87% yield). Notably, cyclic substrates bearing alkyl substituents also provide high levels of coupling efficiency, although the production of minor quantities of regioisomeric adducts can typically be detected due to the presence of secondary sites of hydrogen atom abstraction. Not surprisingly, installation of a *tert*-butyl group on the cyclohexene ring afforded only one major product, with only trace quantities of regioisomers being observed (**17**, 92% yield). Similarly useful levels of regioselectivity were obtained when α-pinene was employed as the olefinic substrate (**18**, 93% yield). The observed selectivity is attributed to the expectation that the electrophilic thiyl radical abstracts the most hydridic hydrogen atom (that is, the hydrogen atom that has the greatest hydride character) and subsequent coupling of the radical proceeds through the least hindered secondary radical position.

Importantly, acyclic alkenes also serve as effective allyl coupling partners in this protocol while generally exhibiting high levels of regiocontrol with respect to the C–C bond forming step (**19**–**23**, 63–90% yield; Fig. 3a). For example, implementation of 2-pentene leads predominantly to the formation of the branched arylation product with excellent efficiency (**20**, 85% yield), with only trace quantities of linear arylation adducts arising from hydrogen atom abstraction of the terminal methylene group. Exposure of 2-methyl-2-pentene to these photoredox conditions afforded two arylation isomers in a 1.4:1 ratio: the major isomer arises from coupling of the allylic radical through the more nucleophilic tertiary substituted carbon, to generate a quaternary centre with useful efficiency (**21**, 90% combined yield for both isomers). In contrast, the reaction with 2,2′-dimethyl-3-hexene yielded only a single product, resulting from coupling at the least hindered terminus of the allylic radical (**22**, 81% yield). With respect to medicinal chemistry applications, it is important to note that heteroatoms can be incorporated into the alkene substrates with no deleterious effect on reaction efficiency (compare **19** and **23**). Thus, coupling of 4-penten-1-ol afforded the desired arylation product in useful yield (**23**, 66% yield). Inclusion of a phenyl substituent on the alkene moiety resulted in a lower isolated yield, presumably owing to decreased hydridicity of the allylic C–H bonds (**24**, 76% yield); however, an analogous substrate bearing a 4-MeO substituent on the phenyl ring underwent efficient, regioselective coupling with 1,4-dicyanobenzene (**25**, 95% yield). Intriguingly, a naturally occurring heteroatom-substituted terpene could be employed directly to generate an arylated derivative in only one step and in good yield (**26**, 84% yield).

This new photoredox–organocatalytic methodology can also be applied to a one-step synthesis of β-aryl ketones. For example, when the trimethylsilyl enol ether of cyclohexanone was exposed to these new arylation conditions, the resultant β-aryl ketone was isolated in high yield (**27**, 83% yield). Furthermore, direct reaction of the corresponding allylic alcohol (2-cyclohexen-1-ol) also furnished ketone **27** via regioselective hydrogen atom abstraction adjacent to the alcohol and subsequent coupling at the allylic β-enol position (see Supplementary Information). Finally, β,γ-unsaturated esters are viable substrates, affording a mixture of β,γ- and γ,δ-unsaturated products (**28**, 81% yield).

We next sought to investigate the scope of the aromatic coupling partner in the arylation protocol. As shown in Fig. 3b, a range of electron-deficient arenes are well tolerated in this new photoredox protocol. Derivatives of dicyanobenzene, including those bearing *ortho*-substituents, readily coupled with the cyclohexenyl allylic radical to afford the corresponding arylation adducts in good to excellent yield (**29**–**31**, 50–83% yield). Moreover, extended aromatic systems, such as a biphenyl derivative, were also found to participate efficiently (**32**, 66% yield). Although the presence of one cyano-substituent was a requirement for the viability of the aryl coupling component, additional electron-withdrawing substituents, such as sulfones, may be readily incorporated without issue (**33**, 69% yield). Notably, a range of cyanopyridine derivatives underwent coupling with cyclohexene to afford the corresponding heteroarylation adducts with useful levels of efficiency (**34**–**37**, 70–84% yield). Additionally, halide substituents at the 2- and 3-positions of pyridine were also tolerated. In these cases it is important to note that the thiol

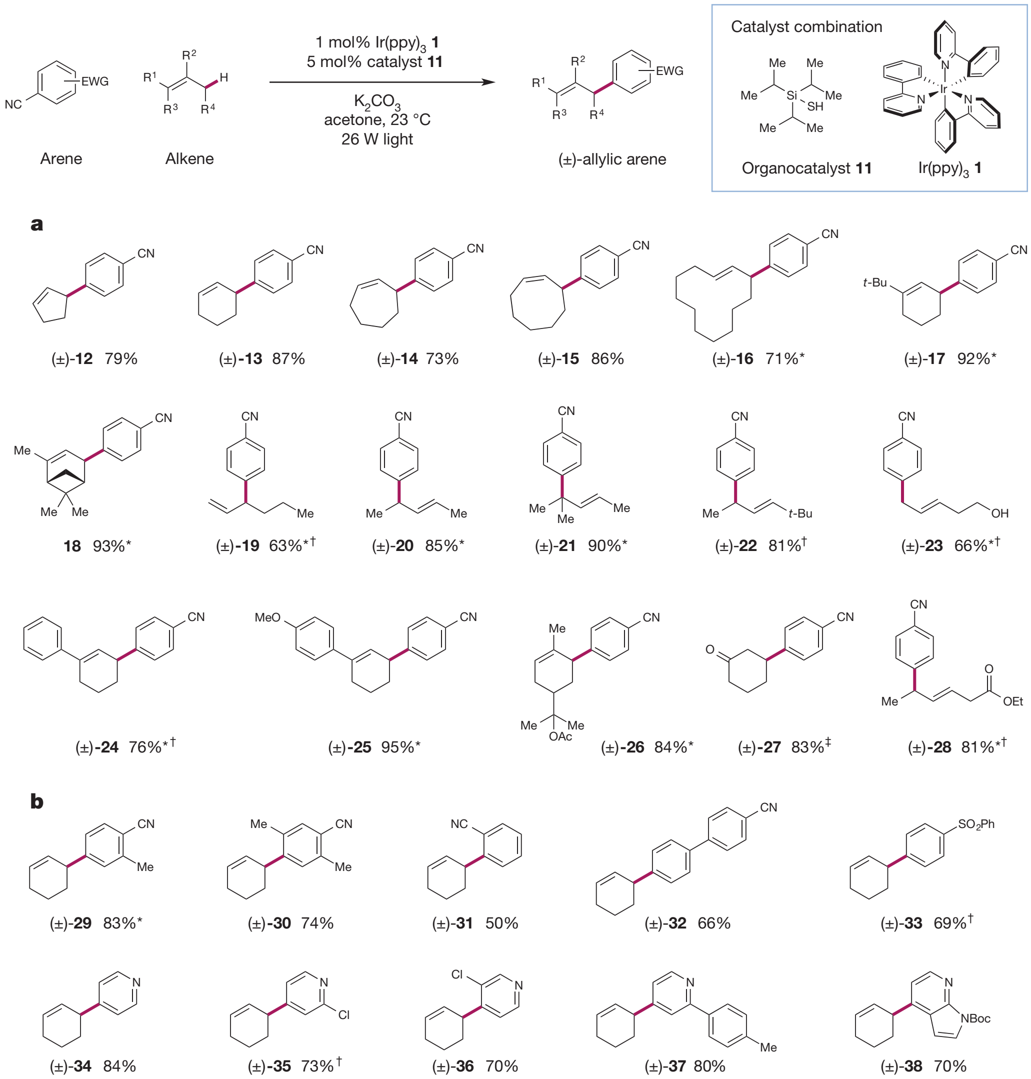


Figure 3 | Substrate scope for the direct allylic arylation reaction. A range of alkenes are efficiently arylated under the standard reaction conditions (top, generalized reaction). EWG, electron-withdrawing group. **a**, The substrate scope includes both cyclic and acyclic alkenes. **b**, A range of arenes bearing electron-withdrawing substituents can be employed as coupling partners under the standard conditions. Isolated yields are indicated below each entry. *Isomers observed; in all cases the major isomer is depicted. Yields refer to the combined yield of all isomers. Ratios of isomers where applicable: (±)-**16** (2.2:1.0 *E*:*Z*), (±)-**17** (>20:1), **18** (>10:1), (±)-**19** (1.4:1.0), (±)-**20** (4.9:1.0), (±)-**21** (1.4:1.0), (±)-**23** (1.1:1.0), (±)-**24** (2.1:1.0), (±)-**25** (>19:1), (±)-**26** (~1:1), (±)-**28** (1.2:1.0), (±)-**29** (1.1:1.0). †Additional thiol or base required; see Supplementary Information for experimental details. ‡Yield from silyl enol ether 83%, yield from 2-cyclohexen-1-ol 62%.

catalyst did not participate in a non-productive pyridine addition S_NAr pathway (an initial concern that did not prove to be valid for the 2-chloropyridine system). Finally, it was shown that a 7-azaindole derivative functions as a viable coupling partner (**38**, 70% yield), a notable finding with respect to potential applications of this technology in medicinal chemistry.

Beyond regiocontrol, a more remarkable selectivity phenomenon that we have observed in these photoredox coupling studies is that mono-arylation adducts are observed exclusively, despite the fact that the allyl substituted aryl product contains a C–H bond that is far weaker than the allylic position of the starting material olefin. Two mechanistic scenarios could account for this unique (and valuable) form of chemoselectivity. First, the doubly activated (benzylic and allylic) C–H bond of the product is significantly less hydridic than the starting material allylic C–H, and is thereby sufficiently less susceptible to hydrogen atom abstraction by the electrophilic thiyl radical of the organocatalyst (thereby governed by the inherent exchange rate constant). Second, it is possible that hydrogen abstraction does indeed occur from the mono-arylation adduct; however, the resultant stabilized radical is insufficiently reactive to participate in a second radical–radical coupling event (a persistent radical effect), and hydrogen atom re-abstraction from solvent or thiol would reconstitute the initial arylation product. Given that we do not observe the formation of any olefin-transposed styrenyl side products, which would arise from the hydrogen atom quenching of a persistent benzylic–allylic radical species, we presume that the former, exchange-constant-gated mechanism is operative.

We next sought to demonstrate the utility of this methodology for the preparation of high-value and versatile building blocks, and at the same time demonstrate the mild nature of the conditions employed (Fig. 4a). More specifically, we have found that boronic ester-substituted cyclohexenes are stable to these photoredox conditions and, more importantly, undergo regioselective arylation at the carbon centre adjacent to the C–B bond (**39**, 73% yield). Moreover, incorporation of a boronic ester substituent on the arene coupling component can also be tolerated in this transformation (**40**, 60% yield).

Finally, a central advantage of this mild, visible-light-mediated arylation protocol is the potential for the late-stage diversification of advanced, highly functionalized synthetic intermediates. As an illustration of this strategy, we subjected 5-pregnen-3β-ol-20-one, a complex biologically active molecule, to our standard reaction conditions. In the event, this compound underwent fragment coupling with 1,4-dicyanobenzene in a highly regioselective and diastereoselective fashion to deliver the aryl functionalized steroid framework **41** in good yield (Fig. 4b).

An overarching goal of this research programme is to define a widely applicable mode of catalytic activation based on well-established physical properties (that is, BDEs, hydrogen atom transfer exchange constants, oxidation potentials), which permits activation of any given substrate in a predictable manner. Although the primary focus of this work has been the direct arylation of allylic C–H bonds, we reasoned that this generic activation mode could be extended to encompass a diverse menu of substrates. In theory, any substrate possessing hydridic C–H bonds of an appropriate strength (~80–90 kcal mol^{-1}) should have the potential

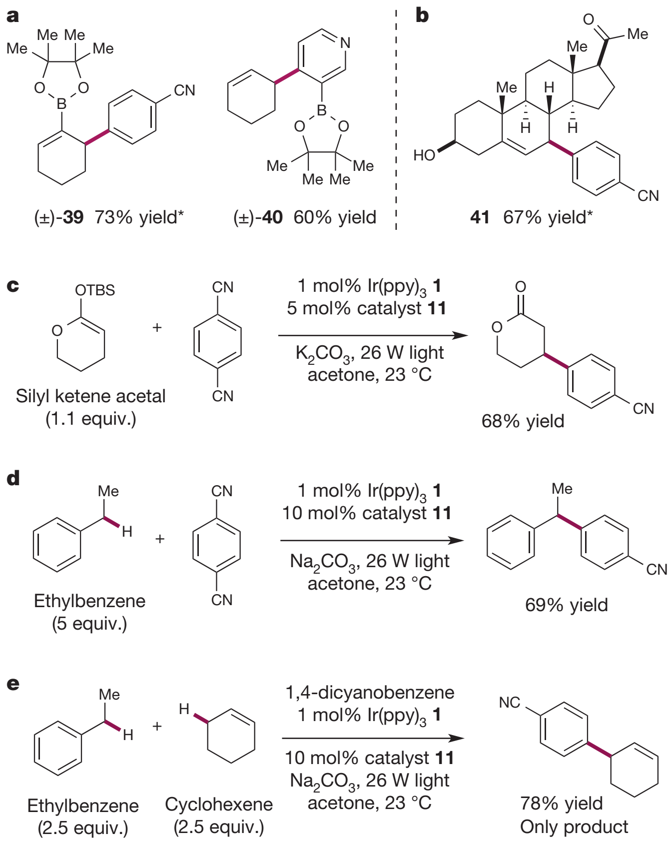


Figure 4 | Expanding the scope of the direct C–H arylation protocol. **a**, Substrates bearing boronic esters substituents are tolerated, providing a means to rapidly access functionalized building blocks. **b**, The mild conditions allow for late-stage functionalization of advanced synthetic intermediates and bioactive natural products. **c**, Silyl ketene acetals are compatible with the reaction conditions, yielding β-aryl lactones. **d**, Arylation is not limited to allylic C–H bonds; benzylic C–H bonds can also be arylated. **e**, The reactivity is governed by bond strengths, with the weaker allylic bond undergoing exclusive functionalization in a direct competition experiment. *Isomers observed; in all cases the major isomer is depicted. Yields refer to the combined yield of all isomers. Ratios of isomers where applicable: (±)-**39** (6:1), **41** (>10:1). See Supplementary Information for experimental details.

to serve as a viable coupling partner in this arylation manifold. Along these lines, we found that silyl ketene acetals are also readily arylated to produce the corresponding β-aryl lactones (Fig. 4c). It is important to consider that lactones represent a high-value substrate class that is incompatible with contemporary enamine mediated β-arylation technologies[13]. Moreover, this fragment coupling protocol can be extended to benzylic substrates such as ethylbenzene to generate the corresponding benzhydryl systems in excellent yields (Fig. 4d).

The predictable and highly useful nature of this hydrogen atom activation mode is exemplified in a direct competition experiment conducted with cyclohexene and ethylbenzene. As shown in Fig. 4e, when both olefinic and benzylic substrates were combined in the same vessel, only the product of allylic arylation was observed (78% yield) with no competitive formation of the benzylic arylation product to any degree. This is a striking result given that ethylbenzene is in fact a suitable substrate for this arylation protocol (see Fig. 4d). The exclusive formation of the allylic arylation product can be readily rationalized by consideration of the BDEs for the two substrates (cyclohexene allylic C–H BDE = 83.2 kcal mol^{-1} versus ethylbenzene benzylic C–H BDE = 85.4 kcal mol^{-1})[15,21], along with an appreciation of the hydridic nature of the respective C–H bonds involved. On this basis, it can be readily anticipated that the thiyl radical will preferentially abstract a hydrogen atom from the allylic substrate, which possesses the weaker and more hydridic C–H bond than the benzylic position. It cannot be ruled out that some degree of benzylic radical formation may occur; however, a rapid hydrogen atom exchange with the cyclohexene substrate may predominate over the rate of productive radical–radical coupling.

We have developed a reaction manifold that permits the direct functionalization and arylation of allylic sp^3 C–H bonds under mild and operationally simple conditions. This new C–C bond forming process, which relies on the mechanistic merger of photoredox and thiol-based organic catalysis, readily accommodates a diverse range of alkene and electron-deficient arene coupling partners. These studies have also established the versatility of this activation mode for the direct arylation of both complex and sensitive olefin-containing molecules.

Received 2 November 2014; accepted 22 January 2015.

1. He, J. *et al.* Ligand-controlled C(sp^3)–H arylation and olefination in synthesis of unnatural chiral α-amino acids. *Science* **343,** 1216–1220 (2014).
2. Eames, J. & Watkinson, M. Catalytic allylic oxidation of alkenes using an asymmetric Kharasch-Sosnovsky reaction. *Angew. Chem. Int. Edn* **40,** 3567–3571 (2001).
3. Johannsen, M. & Jørgensen, K. A. Allylic amination. *Chem. Rev.* **98,** 1689–1708 (1998).
4. Trost, B. M. & Van Vranken, D. L. Asymmetric transition metal-catalyzed allylic alkylations. *Chem. Rev.* **96,** 395–422 (1996).
5. Borg, R. M., Arnold, D. R. & Cameron, T. S. Radical ions in photochemistry. 15. The photosubstitution reaction between dicyanobenzenes and alkyl olefins. *Can. J. Chem.* **62,** 1785–1802 (1984).
6. Hoshikawa, T. & Inoue, M. Photoinduced direct 4-pyridination of C(sp^3)–H bonds. *Chem. Sci.* **4,** 3118–3123 (2013).
7. Sekine, M., Ilies, L. & Nakamura, E. Iron-catalyzed allylic arylation of olefins via C(sp^3)–H activation under mild conditions. *Org. Lett.* **15,** 714–717 (2013).
8. Narayanam, J. M. R. & Stephenson, C. R. J. Visible light photoredox catalysis: applications in organic synthesis. *Chem. Soc. Rev.* **40,** 102–113 (2011).
9. Prier, C. K., Rankic, D. A. & MacMillan, D. W. C. Visible light photoredox catalysis with transition metal complexes: applications in organic synthesis. *Chem. Rev.* **113,** 5322–5363 (2013).
10. Schultz, D. M. & Yoon, T. P. Solar synthesis: prospects in visible light photocatalysis. *Science* **343,** 1239176 (2014).
11. Nicewicz, D. A. & MacMillan, D. W. C. Merging photoredox catalysis with organocatalysis: the direct asymmetric alkylation of aldehydes. *Science* **322,** 77–80 (2008).
12. Nagib, D. A., Scott, M. E. & MacMillan, D. W. C. Enantioselective α-trifluoromethylation of aldehydes via photoredox organocatalysis. *J. Am. Chem. Soc.* **131,** 10875–10877 (2009).
13. Pirnot, M. T., Rankic, D. A., Martin, D. B. C. & MacMillan, D. W. C. Photoredox activation for the direct β-arylation of ketones and aldehydes. *Science* **339,** 1593–1596 (2013).
14. Qvortrup, K., Rankic, D. A. & MacMillan, D. W. C. A general strategy for organocatalytic activation of C–H bonds via photoredox catalysis: direct arylation of benzylic ethers. *J. Am. Chem. Soc.* **136,** 626–629 (2014).
15. Khursan, S. L., Mikhailov, D. A., Yanborisov, V. M. & Borisov, D. I. AM1 Calculations of bond dissociation energies. Allylic and benzylic C–H bonds. *React. Kinet. Catal. Lett.* **61,** 91–95 (1997).
16. Flamigni, L., Barbieri, A., Sabatini, C., Ventura, B. & Barigelletti, F. Photochemistry and photophysics of coordination compounds: iridium. *Top. Curr. Chem.* **281,** 143–203 (2007).
17. Mori, Y., Sakaguchi, Y. & Hayashi, H. Magnetic field effects on chemical reactions of biradical ion pairs in homogeneous fluid solvents. *J. Phys. Chem. A* **104,** 4896–4905 (2000).
18. Ogawa, K. A. & Boydston, A. J. Organocatalyzed anodic oxidation of aldehydes to thioesters. *Org. Lett.* **16,** 1928–1931 (2014).
19. Jencks, W. P. & Salvesen, K. Equilibrium deuterium isotope effects on the ionization of thiol acids. *J. Am. Chem. Soc.* **93,** 4433–4436 (1971).
20. Denisov, E., Chatgilialoglu, C., Shestakov, A. & Denisova, T. Rate constants and transition-state geometry of reactions of alkyl, alkoxyl, and peroxyl radicals with thiols. *Int. J. Chem. Kinet.* **41,** 284–293 (2009).
21. McMillen, D. F. & Golden, D. M. Hydrocarbon bond dissociation energies. *Annu. Rev. Phys. Chem.* **33,** 493–532 (1982).

Supplementary Information is available in the online version of the paper.

Acknowledgements Financial support was provided by NIHGMS (R01 GM103558-03), and we also thank Merck and Amgen for funding. J.D.C. thanks Marie Curie Actions for an International Outgoing Fellowship.

Author Contributions J.D.C. performed and analysed experiments. J.D.C. and D.W.C.M. designed experiments to develop this reaction and probe its utility, and also prepared this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.W.C.M. (dmacmill@princeton.edu).

doi:10.1038/nature14213

Drought impact on forest carbon dynamics and fluxes in Amazonia

Christopher E. Doughty[1], D. B. Metcalfe[2], C. A. J. Girardin[1], F. Farfán Amézquita[3], D. Galiano Cabrera[3], W. Huaraca Huasco[3], J. E. Silva-Espejo[3], A. Araujo-Murakami[4], M. C. da Costa[5], W. Rocha[6], T. R. Feldpausch[7], A. L. M. Mendoza[3], A. C. L. da Costa[5], P. Meir[8,9], O. L. Phillips[10] & Y. Malhi[1]

In 2005 and 2010 the Amazon basin experienced two strong droughts[1], driven by shifts in the tropical hydrological regime[2] possibly associated with global climate change[3], as predicted by some global models[3]. Tree mortality increased after the 2005 drought[4], and regional atmospheric inversion modelling showed basin-wide decreases in CO_2 uptake in 2010 compared with 2011 (ref. 5). But the response of tropical forest carbon cycling to these droughts is not fully understood and there has been no detailed multi-site investigation *in situ*. Here we use several years of data from a network of thirteen 1-ha forest plots spread throughout South America, where each component of net primary production (NPP), autotrophic respiration and heterotrophic respiration is measured separately, to develop a better mechanistic understanding of the impact of the 2010 drought on the Amazon forest. We find that total NPP remained constant throughout the drought. However, towards the end of the drought, autotrophic respiration, especially in roots and stems, declined significantly compared with measurements in 2009 made in the absence of drought, with extended decreases in autotrophic respiration in the three driest plots. In the year after the drought, total NPP remained constant but the allocation of carbon shifted towards canopy NPP and away from fine-root NPP. Both leaf-level and plot-level measurements indicate that severe drought suppresses photosynthesis. Scaling these measurements to the entire Amazon basin with rainfall data, we estimate that drought suppressed Amazon-wide photosynthesis in 2010 by 0.38 petagrams of carbon (0.23–0.53 petagrams of carbon). Overall, we find that during this drought, instead of reducing total NPP, trees prioritized growth by reducing autotrophic respiration that was unrelated to growth. This suggests that trees decrease investment in tissue maintenance and defence, in line with eco-evolutionary theories that trees are competitively disadvantaged in the absence of growth[6]. We propose that weakened maintenance and defence investment may, in turn, cause the increase in post-drought tree mortality observed at our plots.

How does drought affect tropical forests? This question has been studied in long-term experimental drought studies[7,8], in long-term biomass plots that have tracked forest dynamics through drought events[4], and through remote sensing[9–11]. Increased mortality of trees in a large network of 1-ha plot censuses was observed after the 2005 Amazonian drought, turning the forest from an estimated net biomass carbon (C) sink of $\sim$0.71 Mg C ha^{-1} yr^{-1} (ref. 12) to a temporary net source of CO_2 to the atmosphere of twice this, with a total impact (that is, committed source minus baseline sink) of 1.2–1.6 Pg C (ref. 4). An increase in drought-induced tree mortality has also been seen in two multi-year experimentally droughted plots in Amazonia, dominated by an increase in the mortality of large trees[7]. Remote sensing of canopy backscatter after the 2005 drought indicated that, in some parts of Amazonia, the drought caused a change in structure and water content associated with the forest upper canopy. This suggests a slow recovery (>4 years) of forest canopy structure after the severe drought in 2005 (ref. 10).

Because future droughts in tropical regions may increase in frequency and severity[1–3], a better understanding of whether net CO_2 fluxes to the atmosphere from tropical forests increase or decrease during drought periods is urgently required. Drought could either suppress gross primary productivity (GPP), which would lead to an immediate decrease in CO_2 uptake, or it could decrease heterotrophic respiration, thereby decreasing the CO_2 source to the atmosphere, or both[13]. The Amazon basin in 2010 was drier than in 2011, but not warmer, permitting the influences of temperature and precipitation to be separated[5]. A recent atmospheric inversion study in the Amazon basin found that these forests took up 0.25 $\pm$ 0.14 Pg C less CO_2 in 2010 (the year of the drought) than in 2011, after accounting for the effect of increased fires during the drought[5]. A previous study using isotopic techniques found a similar result, with the basin turning from a potential sink to a source after the dry El Niño year of 1997 (ref. 14). These results indicate that annual Amazon droughts apparently suppress photosynthesis more than respiration, but such a relative decrease has not been directly verified with on-the-ground measurements.

To be able to understand and model long-term carbon storage in the tropics, top-down estimates of GPP and fluxes of CO_2 to the atmosphere alone are insufficient. It is also important to understand how the products of photosynthesis are allocated between plant metabolism and biomass growth (NPP) and how that growth is allocated among different organs of the tree[15]. Total autotrophic respiration plus total NPP should approximately equal total GPP over long (multi-year) timescales in a stable environment. However, over shorter timescales the two may differ, because forests may store 'old' carbon in the form of non-structural carbohydrates (NSCs), which may be abundant in tropical forests ($\sim$16 Mg C ha^{-1}, more than enough carbon to rebuild the entire leaf canopy)[16]. These NSCs may function as a reserve that enables the continuation of high rates of growth during periods of lower carbon income from photosynthesis[16–18].

For several years we have measured the main components of total NPP (including 1–3-month records of fine-root, woody and leaf-flush NPP) and autotrophic respiration (including rhizosphere, stem wood and canopy leaf respiration) at thirteen 1-ha rainforest plots in three South American countries, covering contrasting climatic and soil conditions and also across a 2,800 m elevation range in the Andes (Extended Data Tables 1–3). Initial results from these measurements have been described in refs 19–23, presenting complete mean annual sums and mean seasonal cycles of NPP and autotrophic respiration (R_a). This methodology has shown close agreement with independent eddy covariance data on seasonal and annual timescales (Extended Data Fig. 1;

[1]Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford OX1 3QY, UK. [2]Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, 223 62 Lund, Sweden. [3]Universidad Nacional San Antonio Abad de Cusco, Apartado Postal Nro 921, Cusco, Perú. [4]Museo de Historia Natural Noel Kempff Mercado, Universidad Autónoma Gabriel Rene Moreno, Av. Irala 565, Casilla 2489, Santa Cruz, Bolivia. [5]Universidade Federal do Pará, Instituto de Geociências, Faculdade de Meteorologia, Rua Augusto Correa, n° 01, CEP 66075 - 110, Belém, Pará, Brazil. [6]IPAM Instituto de Pesquisa Ambiental da Amazônia Rua Horizontina, 104, Centro, 78640-000 Canarana, Mato Grosso, Brazil. [7]Department of Geography, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4RJ, UK. [8]School of Geosciences, University of Edinburgh, Edinburgh EH9 3FF, UK. [9]Research School of Biology, Australian National University, Canberra, Australian Capital Territory 2601, Australia. [10]School of Geography, University of Leeds, Leeds LS2 9JT, UK.

the slope is within the error of a one-to-one line, 3.0 ± 7.8% (95% confidence interval))[24]. Here we synthesize and further analyse these results to focus specifically on the basin-wide trends before, during and after the 2010 drought, constrained by concurrent measurements in a larger network measuring woody NPP and mortality[4] and inversion studies monitoring changes in atmospheric CO_2 concentrations[5]. Of the thirteen plots, six experienced drought in 2010 (Extended Data Fig. 2). Of these six, three can be considered lowland humid forest more typical of Amazonia (on the basis of species composition and maximum cumulative water deficit (MCWD); see Methods), and three are drier forests at the southern margins of the Amazon forests.

Throughout the two-year period of study, the seven non-drought plots showed steady NPP, R_a and total plant carbon expenditure (PCE, the sum of NPP and R_a, or the carbon expended by the autotrophic metabolism of the ecosystem; Fig. 1, green lines). Total NPP was invariant throughout the drought period at all of our plots (Fig. 1c). Among the six drought-affected plots, there were differences between those in the dry lowlands (Fig. 1, red lines; $n = 3$) and those in the more humid areas (Fig. 1, black lines; $n = 3$). PCE in the humid lowland plots was constant at the start of the drought, but then both PCE and R_a decreased significantly ($P < 0.05$ and $P < 0.01$, respectively; paired t-test, $n = 3$ plots) into early 2011 relative to the 2009 baseline. The humid plots recovered to the 2009 baseline within a few months after the drought, but decreases in R_a at the three dry lowland plots persisted for a year after the 2010 drought (Fig. 1b). This short-term decrease in R_a (dominated by changes in rhizosphere and stem respiration; Extended Data Fig. 7) is in contrast to the results from a multi-annual experimental drought in which R_a increased (dominated by changes in leaf respiration)[19].

PCE should roughly equal total photosynthesis in an ecosystem over annual to multi-annual timescales, with any discrepancy between the two on shorter (monthly) timescales being caused by changes in unmeasured carbon pools such as NSC reserves. A decrease in PCE must therefore equal an equivalent decrease in GPP during a previous period. At our humid drought sites, PCE decreased by 1.90 ± 1.04 Mg C ha^{-1} yr^{-1} (95% confidence interval) during and after the drought period compared with the 2009 baseline (yellow region of Fig. 1a). *In situ* measurements of light-saturated maximum photosynthesis made at a subset of our plots indicate that photosynthesis did decrease significantly ($P < 0.001$, t-test, $n = 20$ trees per plot) during the drought period in comparison with non-drought conditions on the same trees, indicating that the drought was the cause of the decrease in PCE (Extended Data Fig. 3). This measured decrease in photosynthesis is of a similar magnitude to modelled decreases in photosynthesis from drought in eastern Amazonia[25]. We speculate that the asynchrony between the decrease in PCE and the start of the drought indicates that the forests relied on NSC reserves to maintain constant growth and respiration initially during the drought period (Extended Data Fig. 4). Towards the middle of the drought period, R_a decreased in the rhizosphere and stems, whereas NPP and growth continued to remain relatively constant. Because autotrophic respiration consists of maintenance (non-growth) respiration and the respiratory costs associated with growth, this suggests that maintenance respiration not associated with growth must have declined. The decrease in R_a continued after the end of the drought period, potentially allowing the replenishment of the NSC stores once normal photosynthesis resumed (Extended Data Fig. 4). Drought decreased PCE by a larger amount in dry-zone plots than in humid-zone plots, with total PCE continuing to decline into 2011. The greater total decline in PCE is indicative of a larger percentage decrease in total photosynthesis during the drought at the drier plots, a plot-scale observation that matches our *in situ* leaf-level measurements (Extended Data Fig. 3). Our data show little change in net heterotrophic respiration in the humid drought plots (Fig. 1d, black line, and Extended Data Results), and this suggests that the drought forest plots were first a net C source in 2010 due to suppressed photosynthesis, and then a net C sink in early 2011 as photosynthesis returned to normal, while R_a in the stems and rhizosphere remained slightly suppressed in comparison with previous periods (Extended Data Figs 4 and 7).

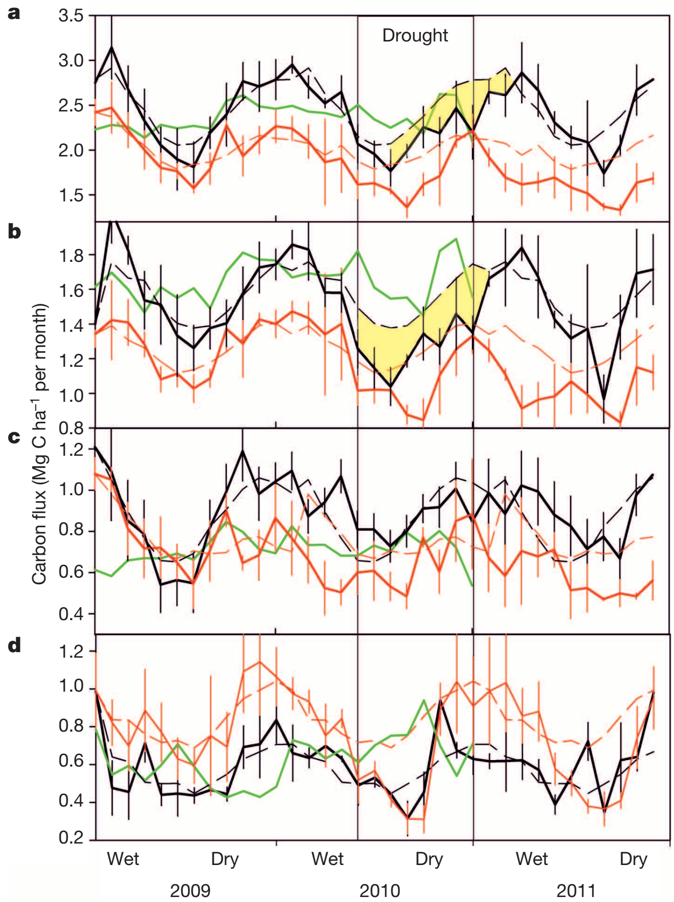


Figure 1 | Impact of drought on carbon fluxes. Results from the forests for the three drought-affected forest plots in humid lowland zones (solid black lines), the three drought-affected forest plots in dry lowland zones (solid red lines), and the seven non-drought plots (solid green lines). **a**, Total PCE; **b**, total R_a; **c**, total NPP; **d**, heterotrophic soil respiration. Error bars indicate the standard error of mean plot differences. For visual clarity we do not include all error bars. Dashed lines show 'normal' (2009; pre-drought) estimates smoothed with a span of 5 months during 2010 and 2011 for the lowland plots (black) and the dry lowland plots (red). The vertical area labelled 'drought' represents the approximate period of the drought. The areas highlighted in yellow represent the drought anomaly or the impact of the drought on total plant carbon expenditure (numerically equivalent to GPP) and R_a.

There was strong seasonality in the components of NPP, with peaks in leaf growth generally anti-correlated with the peaks in woody growth. Hence variation in seasonal growth rates was driven more by shifts in the allocation of NPP than by variation in its total magnitude[26]. NPP allocation in the non-droughted plots did not change significantly between 2009 and 2010 (Fig. 2a–d, green lines). In the droughted plots there were no significant shifts in allocation patterns during the drought period itself, but in the 6 months after the drought there was a significant shift in C allocation for both the humid and dry lowland plots after the drought period away from fine-root growth ($P < 0.01$, paired t-test, $n = 3$) and towards canopy growth (a combination of leaf area index and litterfall (see Methods); $P < 0.05$, paired t-test, $n = 3$) (Fig. 2a–d, red and black lines). Droughts typically increase leaf fall, a strategy that is thought to minimize drought-induced xylem embolisms, and they can also cause temperature-related leaf damage as evaporative cooling decreases[8]. Preferential allocation of carbon towards the canopy in the year after the drought is therefore consistent with known physiological drought responses, and probably represents additional carbon

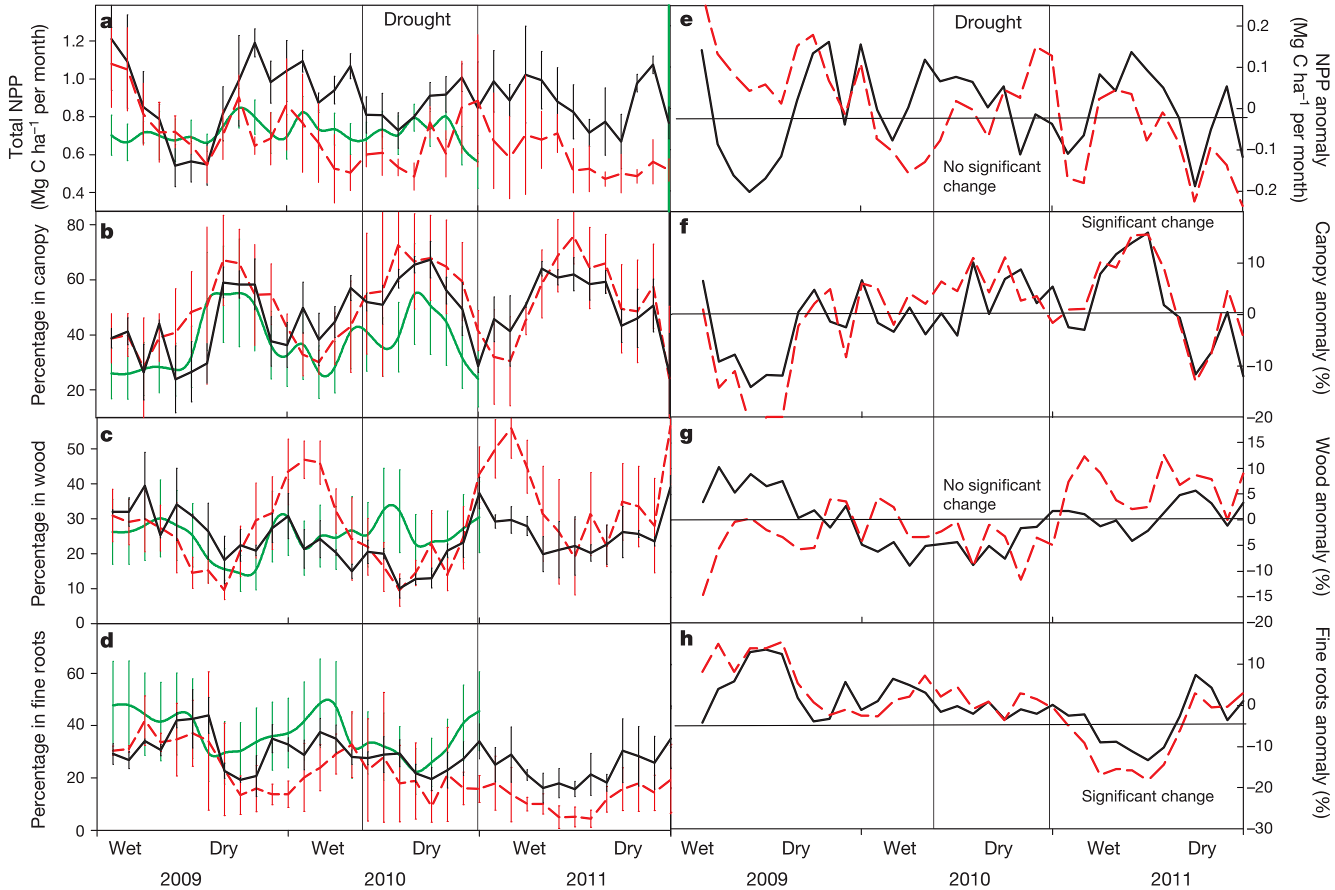


Figure 2 | Impact of drought on carbon allocation. **a**–**d**, Total NPP (**a**), mean carbon allocation to canopy (**b**), mean carbon allocation to wood (**c**) and mean carbon allocation to fine roots (**d**) for non-drought lowland plots (green solid lines; $n = 7$), drought plots in the humid lowlands (black solid lines; $n = 3$) and drought plots in the dry lowlands (red dashed lines; $n = 3$). **e**–**h**, Seasonally detrended anomaly data for each variable on the left. All error bars are standard errors across plots. The vertical bar labelled 'drought' represents the approximate period of the drought. Significant change was determined with a paired *t*-test comparing six-month periods during and after the drought with equivalent months in 2009 for all plots.

to replenish lost and damaged leaves and thereby rebuild photosynthetic capacity. The significant shift away from fine-root growth was surprising, because it has often been assumed that fine-root growth might increase during a drought, but the shift may simply be a reflection of the immediate priority of replacing lost canopy cover instead of a long-term shift away from root growth (for longer-term allocation patterns see Extended Data Fig. 5 and ref. 26).

Individual tree mortality rates roughly doubled at our droughted plots, showing a marginally significant increase ($P = 0.06$, paired one-tailed *t*-test, $n = 5$) from a long-term mean of 1.6 ± 0.6% (Tambopata; $n = 3$) and 2.0 ± 0.4% (Kenia; $n = 2$) to peaks of 3.6% (Tambopata) and 6.7% (Kenia) after the drought (Extended Data Fig. 6c). Mortality remained relatively stable at the non-droughted plots. We tested mortality in a bigger subset of plots at Tambopata and Caxiuanã going back ~30 years at some plots (Extended Data Results) and found that rates of biomass loss increased significantly ($P < 0.05$, Wilcoxon signed-rank test) at Tambopata (drought; $n = 9$) but not at Caxiuanã (no drought; $n = 6$). Committed carbon released as a result of mortality increased from a basin-wide average of 1.6% annual[27] to 3.9% in Tambopata and 2.3% in Kenia (Fig. 3e). Similar drought-induced mortality was also seen across the wider basin after the 2005 drought[4]. The Bolivian plots experienced more severe drought ($MCWD_{anom} < -240$ mm), and here more trees died more quickly than in the Peruvian plots, which were less strongly droughted ($MCWD_{anom} = -51$ mm). Our data indicate that mortality rates peaked 1–2 years after the drought, consistent with the hypothesis that trees were weakened during the drought from decreased maintenance but only succumbed later[19].

Plant carbon expenditure was significantly related ($P < 0.05$, linear regression) and autotrophic respiration was marginally significantly related ($P = 0.08$, linear regression) to the anomaly in MCWD for both annual sums ($n = 13$ individual plots for 2009 minus 2010; $PCE_{anom} = -1.0 + 0.011 \times MCWD_{anom}$, $r^2 = 0.34$, with a standard error on the slope of ±0.004; Fig. 3a, b). The anomaly in NPP, in contrast, showed no significant relationship with the MCWD anomaly (Fig. 3c; $P > 0.10$). We combine a TRMM (version 7; years 1998–2012)-based $MCWD_{anom}$ for each TRMM pixel in the Amazon in 2010 and 2011 with the slope of the above equation (with an intercept of zero) to estimate that mean net total photosynthesis decreased by 0.38 Pg C (0.23–0.53 Pg C) in 2010 compared with 2011, based on a mean South American tropical forested area of 6.77×10^6 km^2 (Fig. 3d). For the same period, an Amazonia-focused atmospheric inversion modelling study estimated a decreased flux of 0.25 ± 0.14 Pg C in 2010 relative to 2011 from decreased photosynthesis, which is within our error estimates[5].

Why would trees prioritize growth over maintenance or defence during and after a drought? This strategy makes sense when viewed from an eco-evolutionary standpoint, where any decrease in growth of an individual tree puts that tree at a competitive disadvantage by increased risk of loss of resources (light, water or nutrients) to neighbours[6]. We speculate that this decrease in maintenance and defence led to our plot-level increase in mortality. Thus, although such a drought-induced strategy may decrease the mean performance per tree in the forest through increased mortality, it is still likely to be selected for on an individual basis given the evolutionary constraints proposed by game theory[28]. In other words, this strategy increases mortality for a small proportion of

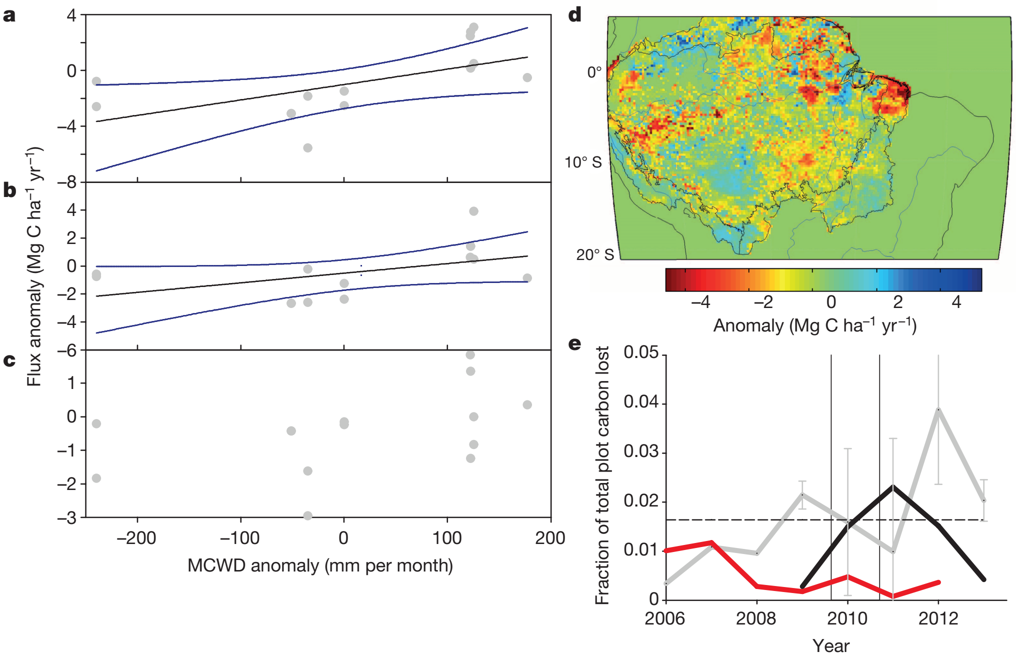


Figure 3 | Estimated impact of drought on the basin-wide flux of CO_2. **a–c**, Shifts in annual fluxes in 2010 relative to 2009 for each individual plot for plant carbon expenditure ($PCE_{anom} = -1.0 + 0.011 \times MCWD_{anom}$, equal to GPP_{anom} over longer timescales; $r^2 = 0.34$, $P < 0.05$) (**a**), R_a ($R_{a,anom} = -0.51 + 0.007 \times MCWD_{anom}$, $r^2 = 0.23$, $P < 0.1$) (**b**) and NPP (not significant) (**c**), plotted against the shift in MCWD in 2010 relative to 2009. **d**, Estimate of basin-wide anomaly in GPP (2011 minus 2010; assumed equal to PCE) based on the TRMM (version 7)-calculated CWD anomaly (mm per month) and the slope of the linear regression found in **a**. We contrast 2010 with 2011 to compare with the atmospheric inversion measurements collected during this period[5]. **e**, Mortality rates as fraction of plot biomass for Peruvian drought plots (grey line; $n = 3$; error bars are standard errors), Bolivian drought plots (black line; $n = 2$) and no-drought plots (red line; $n = 3$). Mean Amazonian background tree mortality (no drought) is shown as a thin black dashed line (from fig. 3 in ref. 27).

trees because most are locked in to growth competition with neighbours. Such unexpected carbon allocation patterns have been theorized previously, but before now they have lacked much empirical support. For instance, trees may grow excess leaves not to improve carbon uptake but to shade out competition[29], or they may over-allocate carbon to root growth in shallow soil systems in response to competition[6].

Overall, our plot data indicate that drought suppressed total CO_2 uptake with little reduction in growth; less carbon was therefore available to the trees for defence and maintenance. Decreased soluble carbon availability may have also increased tree mortality from embolisms and cavitation because NSCs (sugars) may be involved in sensing and reversing embolism[18]. The debate over drought-induced tree mortality is often framed as being caused by carbon starvation, water cavitation or biotic attack, but the three are often intertwined[30] because during drought there is less carbon available to fend off all three threats. This insight and new mechanistic understanding can help to improve predictions of the impact of future climate change on tropical forests.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 February; accepted 22 December 2014.

1. Lewis, S. L., Brando, P. M., Phillips, O. L., van der Heijden, G. M. F. & Nepstad, D. The 2010 Amazon drought. *Science* **331,** 554 (2011).
2. Gloor, M. *et al.* Intensification of the Amazon hydrological cycle over the last two decades. *Geophys. Res. Lett.* **40,** 1729–1733 (2013).
3. Solomon, S. *et al.* (eds) *Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, 2007).
4. Phillips, O. L. *et al.* Drought sensitivity of the Amazon rainforest. *Science* **323,** 1344–1347 (2009).
5. Gatti, L. V. *et al.* Drought sensitivity of Amazonian carbon balance revealed by atmospheric measurements. *Nature* **506,** 76–80 (2014).
6. Franklin, O. *et al.* Modeling carbon allocation in trees: a search for principles. *Tree Physiol.* **32,** 648–666 (2012).
7. da Costa, A. C. L. *et al.* Effect of 7 yr of experimental drought on vegetation dynamics and biomass storage of an eastern Amazonian rainforest. *New Phytol.* **187,** 579–591 (2010).
8. Nepstad, D. C. *et al.* The effects of partial throughfall exclusion on canopy processes, aboveground production, and biogeochemistry of an Amazon forest. *J. Geophys. Res.* **107** (D20), 8085 (2002).
9. Xu, L. A. *et al.* Widespread decline in greenness of Amazonian vegetation due to the 2010 drought. *Geophys. Res. Lett.* **38,** L07402 (2011).
10. Saatchi, S. *et al.* Persistent effects of a severe drought on Amazonian forest canopy. *Proc. Natl Acad. Sci. USA* **110,** 565–570 (2013).
11. Saleska, S. R., Didan, K., Huete, A. R. & da Rocha, H. R. Amazon forests green-up during 2005 drought. *Science* **318,** 612 (2007).
12. Phillips, O. L. *et al.* Changes in the carbon balance of tropical forests: evidence from long-term plots. *Science* **282,** 439–442 (1998).
13. Meir, P., Metcalfe, D. B., Costa, A. C. L. & Fisher, R. A. The fate of assimilated carbon during drought: impacts on respiration in Amazon rainforests. *Phil. Trans. R. Soc. B* **363,** 1849–1855 (2008).
14. Townsend, A. R., Asner, G. P., White, J. W. C. & Tans, P. P. Land use effects on atmospheric C-13 imply a sizable terrestrial CO_2 sink in tropical latitudes. *Geophys. Res. Lett.* **29,** 1426 (2002).
15. Malhi, Y., Doughty, C. & Galbraith, D. The allocation of ecosystem net primary productivity in tropical forests. *Phil. Trans. R. Soc. B* **366,** 3225–3245 (2011).
16. Wurth, M. K. R., Pelaez-Riedl, S., Wright, S. J. & Korner, C. Non-structural carbohydrate pools in a tropical forest. *Oecologia* **143,** 11–24 (2005).
17. Dietze, M. C. *et al.* Nonstructural carbon in woody plants. *Annu. Rev. Plant Biol.* **65,** 667–687 (2014).
18. Sala, A., Woodruff, D. R. & Meinzer, F. C. Carbon dynamics in trees: feast or famine? *Tree Physiol.* **32,** 764–775 (2012).
19. da Costa, A. C. L. *et al.* Ecosystem respiration and net primary productivity after 8–10 years of experimental through-fall reduction in an eastern Amazon forest. *Plant Ecol. Divers.* **7,** 7–24 (2014).
20. Araujo-Murakami, A. *et al.* The productivity, allocation and cycling of carbon in forests at the dry margin of the Amazon forest in Bolivia. *Plant Ecol. Divers.* **7,** 55–69 (2014).
21. Doughty, C. E. *et al.* The production, allocation and cycling of carbon in a forest on fertile terra preta soil in eastern Amazonia compared with a forest on adjacent infertile soil. *Plant Ecol. Divers.* **7,** 41–53 (2014).
22. Malhi, Y. *et al.* The productivity, metabolism and carbon cycle of two lowland tropical forest plots in south-western Amazonia, Peru. *Plant Ecol. Divers.* **7,** 85–105 (2014).
23. Rocha, W. *et al.* Ecosystem productivity and carbon cycling in intact and annually burnt forest at the dry southern limit of the Amazon rainforest (Mato Grosso, Brazil). *Plant Ecol. Divers.* **7,** 25–40 (2014).

24. Fenn, K., Malhi, Y., Morecroft, M., Lloyd, C. & Thomas, M. The carbon cycle of a maritime ancient temperate broadleaved woodland at seasonal and annual scales. *Ecosystems* http://dx.doi.org/10.1007/s10021-014-9793-1 (13 December 2014).
25. Fisher, R. A. *et al.* The response of an Eastern Amazonian rain forest to drought stress: results and modelling analyses from a throughfall exclusion experiment. *Glob. Change Biol.* **13,** 2361–2378 (2007).
26. Doughty, C. E. *et al.* Allocation trade-offs dominate the response of tropical forest growth to seasonal and interannual drought. *Ecology* **95,** 2192–2201 (2014).
27. Lewis, S. L. *et al.* Concerted changes in tropical forest structure and dynamics: evidence from 50 South American long-term plots. *Phil. Trans. R. Soc. B* **359,** 421–436 (2004).
28. King, D. A. A model analysis of the influence of root and foliage allocation on forest production and competition between trees. *Tree Physiol.* **12,** 119–135 (1993).
29. Hikosaka, K. & Anten, N. P. R. An evolutionary game of leaf dynamics and its consequences for canopy structure. *Funct. Ecol.* **26,** 1024–1032 (2012).
30. McDowell, N. G. *et al.* The interdependence of mechanisms underlying climate-driven vegetation mortality. *Trends Ecol. Evol.* **26,** 523–532 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Brando and Tanguro partners for logistical support and advice. This work is a product of the Global Ecosystems Monitoring (GEM) network (http://gem.tropicalforests.ox.ac.uk) and the RAINFOR and ABERG research consortia, and was funded by grants to Y.M. and O.L.P. from the Gordon and Betty Moore Foundation to the Amazon Forest Inventory Network (RAINFOR) and the Andes Biodiversity and Ecosystems Research Group (ABERG), and grants from the UK Natural Environment Research Council (NE/D01025X/1, NE/D014174/1, NE/F002149/1 and NE/J011002/1), the NERC AMAZONICA consortium grant (NE/F005776/1) and the EU FP7 Amazalert (282664) GEOCARBON (283080) projects. Some data in this publication were provided by the Tropical Ecology Assessment and Monitoring (TEAM) Network, a collaboration between Conservation International, the Missouri Botanical Garden, the Smithsonian Institution and the Wildlife Conservation Society, and partly funded by these institutions, the Gordon and Betty Moore Foundation, and other donors. T.R.F. is supported by a National Council for Scientific and Technological Development (CNPq, Brazil) award. P.M. is supported by an ARC fellowship award FT110100457; O.L.P. is supported by an ERC Advanced Investigator Award and a Royal Society Wolfson Research Merit Award; Y.M. is supported by an ERC Advanced Investigator Award and by the Jackson Foundation. C.E.D. acknowledges funding from the John Fell Fund.

Author Contributions C.E.D., Y.M. and D.B.M. designed and implemented the study. Y.M. conceived the GEM network, C.E.D., D.B.M., C.A.J.G. and Y.M. implemented it, and O.L.P. contributed to its development. C.E.D. analysed the data. C.E.D., C.A.J.G., F.F.A., D.G., W.H.H., J.E.S., A.A., M.C.C., A.C.L.C., T.F., A.M., W.R. and O.L.P. collected the data. C.E.D. wrote the paper with contributions from Y.M., O.L.P., P.M. and D.B.M.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.E.D. (chris.doughty@ouce.ox.ac.uk).

METHODS

We measured total NPP and autotrophic respiration at thirteen 1-ha plots (plots described individually below) throughout the Amazon basin in the period 2009–2010 (and 2009–2011 or 2009–2012 for droughted plots). A detailed description of each measurement is listed in Extended Data Tables 1–3. Total measured NPP included canopy, woody, and fine-root NPP. In our seasonal estimates of NPP we exclude several smaller components such as branch fall (although branch-fall data are shown in Extended Data Fig. 6), herbivory, coarse root and small-tree (<10 cm) NPP that we have included in previous estimates of these sites. We calculate leaf flush by calculating the change in leaf area index, LAI ($m^2\,m^{-2}$), multiplied by the mean specific leaf area ($m^2\,g^{-1}$), and adding this to leaf litterfall by following a procedure from ref. 31. Total estimated autotrophic respiration consisted of rhizosphere respiration (that is, respiration from roots, mycorrhizae and exudate-dependent soil microbes), woody respiration and canopy respiration. Each component was measured every 1–3 months, except for canopy respiration, which was measured only 1–2 times per plot at the leaf level but scaled to the canopy scale using monthly LAI partitioned into sun and shade components. Seasonal changes in autotrophic respiration during and following drought are due to monthly measured rhizosphere and woody tissue (stem) respiration, not canopy (leaf) respiration (Extended Data Fig. 7). Detailed information on the methodology and graphs showing data from each individual component are also available from a series of companion papers[19–23,32,33]. Each of these site papers includes a full spatial and scaling error analysis for each measurement, so, for brevity, we do not include them here. All raw data inputs are available on request from the authors or from http://gem.tropicalforests.ox.ac.uk/.

Photosynthesis. Light-saturated leaf photosynthesis was measured at two plots in Bolivia in the peak of the drought (November 2010) and during a non-drought period (June 2011) on the same ~20 individual trees (12 different species from plot A and 17 species from plot B) in the plot, using canopy-top cut branches (immediately recut under water to restore hydraulic conductivity). These measurements are compared with leaf photosynthesis measurements in the Tapajos, Brazil, on attached (not cut) canopy top leaves accessed via three walk-up towers, to test whether A_{sat} (light-saturated photosynthesis) would be expected to decrease during a typical (non-severe) dry season, and the measurements were taken at the start of a typical dry season to the end of the dry season (Extended Data Fig. 3; methodological details in Extended Data Tables 1–3).

Climate. We classified our drought sites according to anomalies in cumulative water deficit (CWD) based on precipitation data collected from automatic weather stations at each of the plots (Skye Instruments). Six of our 13 plots experienced drought in 2010 (negative CWD anomalies for more than half the year), with a mean CWD anomaly of −107 mm in October and a mean maximum CWD (MCWD) of −135 mm, meaning that the driest month on average had a water deficit 135 mm greater than in a normal year (Extended Data Fig. 2). This varied regionally: the highest MCWD was in the Bolivian sites ($MCWD_{anom} = -240$ mm) and the lowest was in the lowland Peruvian sites ($MCWD_{anom} = -51$ mm). We used Tropical Rainfall Monitoring Mission (TRMM) data from January 1998 to December 2012 (TRMM version 7) to calculate for each pixel the maximum monthly CWD anomaly (Extended Data Fig. 2). The basin-wide median $MCWD_{anom}$ for 2010 for droughted tropical forest regions was 136 mm (excluding $MCWD_{anom} \geq 0$ mm). This implies that the mean of our droughted plots had a moisture anomaly equivalent to the basin-wide 'typical' Amazon drought for 2010 (Extended Data Fig. 2), but also that our plots did not experience the more severe drought seen by some regions of Amazonia.

Statistics. All data were tested for normality, and if they were normal we did a two-tailed paired *t*-test using Sigmaplot (Systat Software Inc.). If normality was not passed, as with the mortality data, we used a Wilcoxon signed-rank test. We used a two-tailed test except for mortality, for which we expected the change to be in one direction and therefore used a one-tailed test. We calculated 95% confidence intervals by multiplying the standard error by 1.96.

Additional mortality data. For the additional RAINFOR analyses for Tambopata and Caxiuanã, interval-by-interval loss rates in each plot were computed by following standard RAINFOR field and ForestPlots.net data protocols (see, for example, refs 34, 35). At Caxiuanã, data were collected by the TEAM network, whose protocols are closely based on RAINFOR models. These include multiple repeated diameter measurements of the same tree at 1.3 m or above buttresses (allowing where necessary for changes in point of measurement), high-resolution botanical identifications of hundreds of tree species at each site, and the use of taxon-specific wood density values, to derive from each individual tree at least 10 cm in diameter the stand-level values of biomass and biomass dynamics. We used a generalized region-specific height–diameter biomass allometry[36]. Because here the question is simply whether the 2010 drought coincided with mortality changes in each site, and not what the precise values of mortality were for individual intervals and plots, we did not attempt to account for the small effects of slightly varying census-interval lengths on wood production rates. Data were downloaded from ForestPlots.net in October 2014 and from the TEAM database in April 2013.

Site descriptions of thirteen 1-ha plots. Plots with drought in 2010. The Kenia plots ($n = 2$, 1-ha plots) were established and monitored on private property at the Hacienda Kenia in Guarayos Province, Santa Cruz, Bolivia (16.0158° S, 62.7301° W) from January 2009. The plots are 2 km apart and are situated on inceptisols with relatively high fertility (high cation exchange capacity and phosphorus concentration) and low acidity compared with eastern Amazonian forests. The plots experienced almost identical climate and had sandy loam soil with 76% sand content. However, one plot was located on a shallow soil (<1 m depth) over pre-Cambrian bedrock, leading to lower available water (we term this plot Kenia-B). The second plot was located on deeper soils in a slight topographic depression (henceforth termed Kenia-A). These differences in drainage and soil depth had an effect on forest composition at this ecotone, with Kenia-A hosting a more humid forest type typical of Amazonian forests and Kenia-B a drier forest type typical of *chiquitano* dry forests. For further details see ref. 20.

The Tanguro plots ($n = 2$, 1-ha plots) are located on the Fazenda Tanguro (~80,000 ha) in Mato Grosso state, about 30 km north of the southern boundary of the Amazon rainforest in Brazil (13.0765° S, 52.3858° W). The soil type at the site is a red–yellow alic dystrophic latosol (RADAM Brazil, 1974; Brazilian soil classification), a relatively infertile sandy ferralsol (FAO classification) or oxisol (Haplustox; US Department of Agriculture classification scheme), the groundwater is at ~15 m depth, and no layers of soil prevent root penetration through the soil profile. These soils are among the least fertile in Amazonia and are widespread across eastern Amazonia. The vegetation is closed-canopy, old-growth forest with a relatively low mean canopy height (20 m) and relatively low plant species diversity (97 species of trees and lianas greater than 10 cm DBH (diameter at 1.3 m stem height above the ground)) when compared with the wetter forests typical of the central Amazon. For further details see ref. 23.

The Tambopata study plots ($n = 2$, 1-ha plots) are located in the Tambopata reserve (TAM-05, 12.837° S, 69.2937° W; TAM-06, 12.828° S, 69.2690° W), in the Madre de Dios region of Peru. The geomorphology of the study region is based on old floodplains of the meandering Tambopata River. TAM-05 is situated on a Pleistocene terrace (<100 kyr old). The soil at TAM-05 is a haplic cambisol (WRB taxonomy), and that at TAM-06 is a haplic alisol[37]. We incorporate mortality data from an additional nearby plot (TAM-09). No hardpan layers of soil prevent root penetration through the soil profile. For further details see ref. 22.

We divided these six plots into three lowland plots (TAM-05, TAM-06 and Kenia-A; black lines in Figs 1 and 2) and three dry-lowland plots (the two Tanguro plots and Kenia-B; red lines in Figs 1 and 2). Distinction of dry-lowland plots was made by using mean MCWD for Tanguro and by species composition for Kenia-B with drier forest type species typical of *chiquitano* dry forests.

Plots with no drought in 2010. The San Pedro site ($n = 2$, 1-ha plots) at 13.0491° S, 71.5365° W is located in the Kosñipata Valley, in the cultural buffer zone of the Parque Nacional del Manú, Cusco, Peru. The two plots at San Pedro lie very close to the transition between upper and pre-montane forest zones, which occurs in this valley at ~1,500–2,000 m. Although data on cloud cover frequency and cloud base elevation in the plots over the annual cycle are difficult to obtain, SP 1750 is immersed for longer periods than SP 1500 during the austral winter months. SP 1500 is estimated to be near the lower limit of the cloud base. For further details see ref. 33.

The Wayqecha and Esperanza plots ($n = 2$, 1-ha plots) (Wayqecha, RAINFOR plot code WAY-01, 13.1751° S 71.5948° W; Esperanza, RAINFOR plot code ESP-01) plots are high-elevation cloud forest located in the cultural buffer zone of the Parque Nacional del Manú, Cusco, Peru at ~3,000 m elevation. The two plots lay a few hundred metres below the treeline transition to high-elevation grasslands. For further details see ref. 32.

The Caxiuanã plots (CAX-08 and CAX-06; $n = 2$, 1-ha plots) are located in Caxiuanã National Forest Reserve, Pará, in the eastern Brazilian Amazon. The *terra preta* plot (1.8560° S, 51.4352° W; plot code CAX-08 in the RAINFOR Amazon forest inventory network) was a late successional forest with a large proportion of fruit trees, on an isolated patch (<2 ha) of fertile dark earth or *terra preta do Indio*. The original ferralsol soils became progressively enriched by the activities of local inhabitants between AD 1280 and 1600 (ref. 38). The species composition of the *terra preta* plot was that of an old abandoned agroforestry system, with Brazil nut (*Bertholletia excelsa*) and kapok (*Ceiba pentandra*) as well as palaeotropical tree crops including coffee (*Coffea*) and orange (*Citrus*). The water-side location of the *terra preta* plot results in a substantially different microclimate from that of the inland tower plot, with high solar radiation (the large cool water area of the bay suppresses cloud formation close to the bay) and higher temperatures. The 'tower' plot (CAX-06; 1.7198° S, 51.4581° W) was a tall primary forest (35 m canopy height) situated on a clay-rich geric alumnic ferralsol (alumnic, hyperdistric, clayic) near an

eddy covariance flux tower, with species composition typical of eastern Amazonia. For further details see ref. 21.

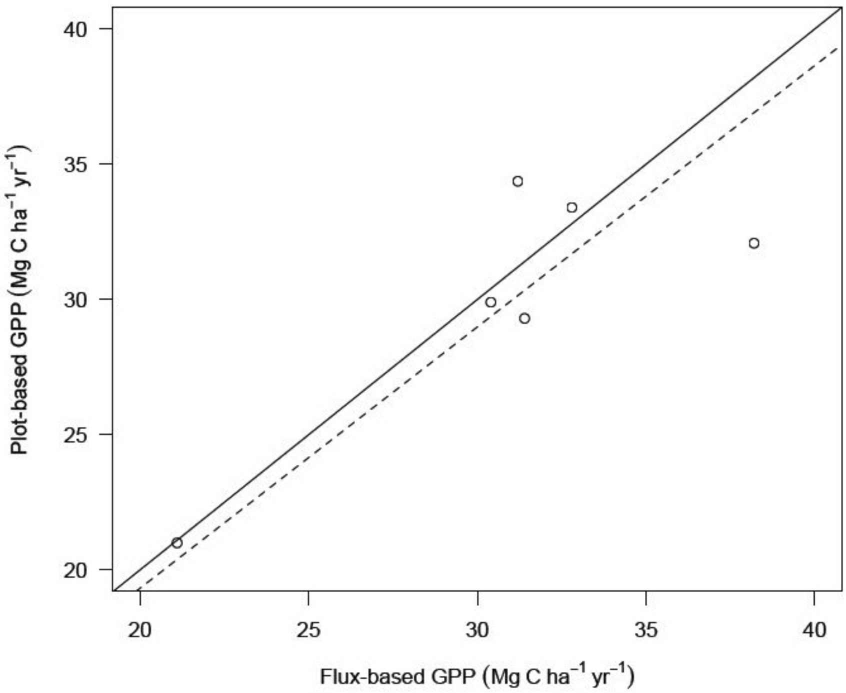
The Caxiuanã plot (TFE-control; $n = 1$, 1-ha plot) is the control plot of an experimental drought study which is ~2 km south of the tower plot mentioned above (1.7279° S, 51.4680° W). It is a largely undisturbed *terra firme* forest, of the type widespread across eastern Amazonia. The study plot is located on highly weathered vetic acrisols typical of upland forests in the eastern Amazon, with a thick stony laterite layer at 3–4 m depth. The site elevation is 15 m above river level in the dry season, and the water table has occasionally been observed at a soil depth of 10 m during the wet season. For further details see ref. 19.

Extended Data results. Heterotrophic respiration. Soil heterotrophic respiration showed no significant change during the drought period in the droughted humid lowland plots (Fig. 1d, black line; $n = 3$) and no significant change with CWD anomaly ($P > 0.05$). There was a slight suppression of heterotrophic soil respiration near the start of the drought, but this was compensated for by a larger than normal increase in heterotrophic soil respiration later in the drought as some rains (although much lower than normal) arrived (Fig. 1d, black line). However, in contrast, the droughted dry lowland plots did show a large decrease in soil heterotrophic respiration at the start of the drought in comparison with 2009 (although only marginally significant: $P < 0.1$, $n = 3$) (Fig. 1d, red line), but these regions are a geographically small part of the basin and their overall influence on basin wide fluxes is likely to be small. Mean temperatures were similar in 2010 and 2011 and therefore any change in heterotrophic flux was most likely to have been driven by moisture (Extended Data Fig. 2). Deadwood respiration was initially suppressed during the dry season of the drought year, but this was compensated for by a large gain once the rains started, leading to no net annual change in deadwood respiration from the drought (Extended Data Fig. 6). Branch fall did not increase during the drought and, in fact, slightly decreased, possibly because of lower wind speeds from decreased storm activity (Extended Data Fig. 6). Our data show little net change in heterotrophic respiration; we therefore estimate that the drought forest plots were first a net C source in 2010 as a result of suppressed photosynthesis, and then a net C sink in early 2011 as photosynthesis returned to normal but R_a remained slightly supressed compared with previous periods, an observation in line with a recent atmospheric inversion study of the Amazon basin[5].

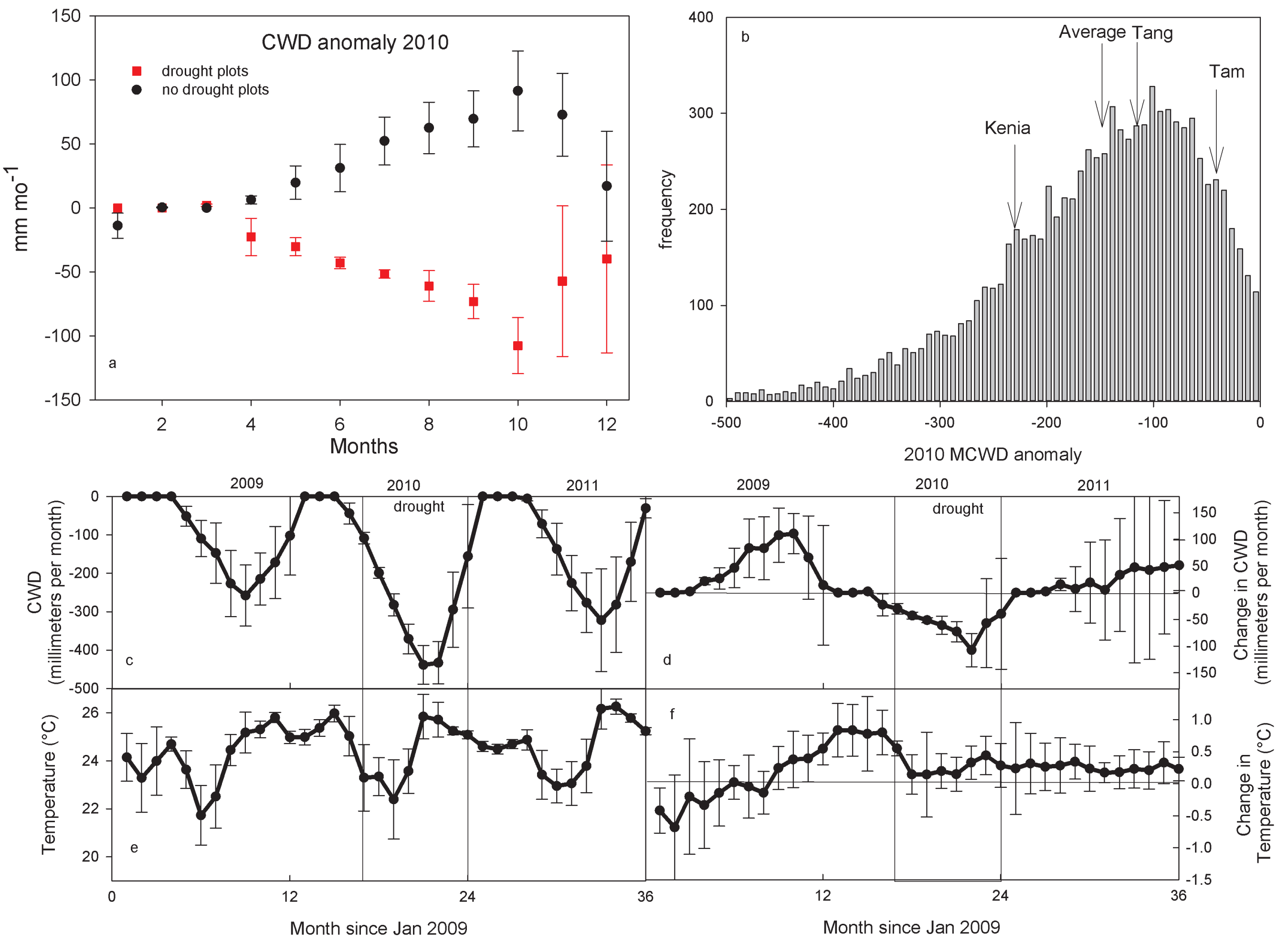
Shifts in carbon allocation. In two of the plots (Kenia-A and Kenia-B), NPP allocation shifted towards roots in the second year after the drought, possibly to alleviate water stress for future droughts, or to increase nutrient uptake to track recovered carbon uptake (Extended Data Fig. 5; NPP allocation patterns at this site are explored in detail in ref. 26). However, allocation responses to drought vary strongly by site. For instance, in two lowland Peruvian plots that experienced milder drought, NPP instead shifted back towards woody growth in the second year after the drought (Extended Data Fig. 5), whereas in two dry lowland Brazilian plots that experienced moderate drought, woody growth increased in the year after the drought at the expense of canopy and fine-root growth (Extended Data Fig. 5). The two plots hardest hit by the drought ($MCWD_{anom} = -240$ mm) showed a long-term decrease in allocation of NPP towards wood even though total NPP remained constant (Extended Data Fig. 5). This indicates that care should be taken in the interpretation of tree growth and dendrochronology results as proxies for productivity after drought, because they may be influenced more by shifting carbon allocation than by changes in total NPP. Our plots showed no significant change in woody NPP growth rates during the drought, although there was a small decline (Fig. 2g). Woody growth rates may actually have declined, but our sample size of three was too small to capture the signal statistically.

Additional mortality results. To see whether mortality increased more broadly in the regions surrounding our plots, we compared plots in the RAINFOR database near Tambopata (with drought according to our meteorological station data) with Caxiuanã (without drought in 2010). In Caxiuanã, we compared plots 1–6 (TEC-01 to TEC-06, using the RAINFOR code) for pre-2010 mortality (starting in 2003) with mortality from a census in late 2010. In Tambopata, we compared plots TAM-01 to TAM-08 for pre-2010 mortality (mostly starting in 1983) with mortality from a census in mid-2011. For this data set, we used a non-parametric Wilcoxon signed-rank test and found significant increase in biomass mortality following the 2010 drought in the larger Tambopata data set ($n = 9$, $P = 0.018$). This is in contrast with Caxiuanã (a no-drought site), where we also had high-resolution meteorological station data, and found no significant change after 2009 ($n = 6$, $P > 0.05$).

31. Doughty, C. E. & Goulden, M. L. Seasonal patterns of tropical forest leaf area index and CO_2 exchange. *J. Geophys. Res.* **113,** G00b06 (2008).
32. Girardin, C. A. J. *et al.* Productivity and carbon allocation in a tropical montane cloud forest in the Peruvian Andes. *Plant Ecol. Divers.* **7,** 107–123 (2014).
33. Huasco, W. H. *et al.* Seasonal production, allocation and cycling of carbon in two mid-elevation tropical montane forest plots in the Peruvian Andes. *Plant Ecol. Divers.* **7,** 125–142 (2014).
34. Quesada, C. A. *et al.* Basin-wide variations in Amazon forest structure and function are mediated by both soils and climate. *Biogeosciences* **9,** 2203–2246 (2012).
35. Lopez-Gonzalez, G., Lewis, S. L., Burkitt, M. & Phillips, O. L. ForestPlots.net: a web application and research tool to manage and analyse tropical forest plot data. *J. Veg. Sci.* **22,** 610–613 (2011).
36. Feldpausch, T. R. *et al.* Tree height integrated into pantropical forest biomass estimates. *Biogeosciences* **9,** 3381–3403 (2012).
37. Quesada, C. A. *et al.* Soils of Amazonia with particular reference to the RAINFOR sites. *Biogeosciences* **8,** 1415–1440 (2011).
38. Lehmann, J., Kern, D. C., Glaser, B. & Woods, W. I. *Amazonian Dark Earths: Origin, Properties, Management* (Kluwer Academic, 2003).
39. Malhi, Y. *et al.* The linkages between photosynthesis, productivity, growth and biomass in lowland Amazonian forests. *Global Change Biol.* http://dx.doi.org/10.1111/gcb.12859 (10 January 2015).
40. Malhi, Y. *et al.* Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest. *Proc. Natl Acad. Sci. USA* **106,** 20610–20615 (2009).
41. Doughty, C. E. An *in situ* leaf and branch warming experiment in the Amazon. *Biotropica* **43,** 658–665 (2011).
42. da Rocha, H. R. *et al.* Seasonality of water and heat fluxes over a tropical forest in eastern Amazonia. *Ecol. Appl.* **14,** S22–S32 (2004).
43. Brando, P. M. *et al.* Abrupt increases in Amazonian tree mortality due to drought–fire interactions. *Proc. Natl Acad. Sci. USA* **111,** 6347–6352 (2014).
44. Chave, J. *et al.* Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* **145,** 87–99 (2005).
45. Martin, A. R. & Thomas, S. C. A reassessment of carbon content in tropical trees. *PLoS ONE* **6,** e23533 (2011).
46. Metcalfe, D. B. *et al.* Factors controlling spatio-temporal variation in carbon dioxide efflux from surface litter, roots, and soil organic matter at four rain forest sites in the eastern Amazon. *J. Geophys. Res.* **112,** G04001 (2007).
47. Malhi, Y. *et al.* Comprehensive assessment of carbon productivity, allocation and storage in three Amazonian forests. *Global Change Biol.* **15,** 1255–1274 (2009).
48. Chambers, J. Q. *et al.* Respiration from a tropical forest ecosystem: Partitioning of sources and low carbon use efficiency. *Ecol. Appl.* **14,** S72–S88 (2004).

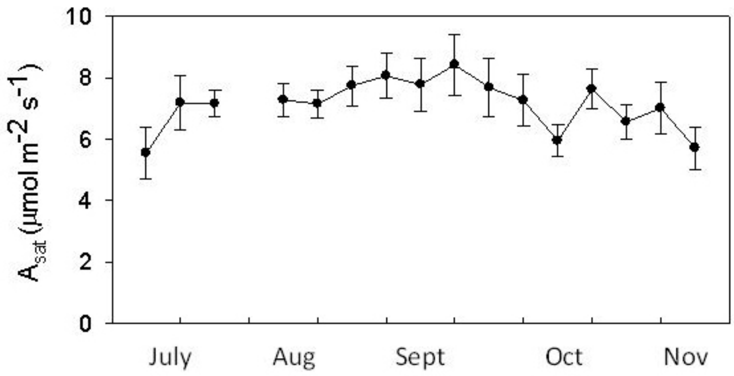


Extended Data Figure 1 | Comparison of plot-based and flux-based estimates of gross primary productivity (GPP) at six different sites worldwide. The black line represents 1:1; the dashed line represents a linear fit with a *y* intercept of 0. Slope = 0.97 ± 0.04, coefficient of determination = 0.61. If the Caxiuanã tower site is removed, then slope = 1.01 ± 0.03 and coefficient of determination = 0.87. Data points are from Manaus, Tapajos, Caxiuanã (Brazil), Wytham Woods (UK) and Lambir Hills (Malaysia). For further details see ref. 39.

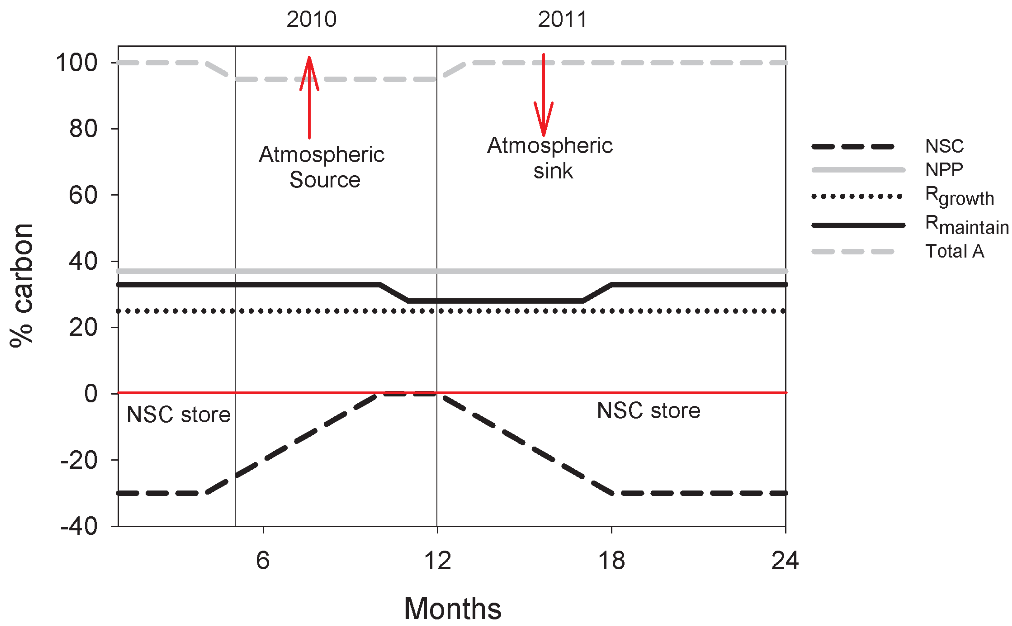


Extended Data Figure 2 | Climate data for the plots. **a**, Cumulative water deficit (CWD) anomaly for six droughted plots (red) and the remaining seven non-droughted plots (black) in 2010, on the basis of data from Skye instruments meteorological stations near each plot. Meteorology stations were set up in either ~2005 ($n = 4$) or ~2009 ($n = 4$). **b**, MCWD anomaly for 2010 (the minimum of CWD_{mean} minus CWD_{2010}) for the entire Amazon basin based on TRMM version 7 data (1998–2012). For clarity we do not show MCWD for non-droughted sites for which $MCWD_{anom} = 0$; this is the maximum potential value because by definition the wettest average month has a CWD value of 0). The arrows depict the site-specific MCWD anomaly for each drought area and average for all drought plots. **c**, **e**, Data from Skye instruments meteorological stations from January 2009 to December 2011 near our drought plots for cumulative water deficit (mm per month) (**c**) and air temperature (°C) (**e**). **d**, **f**, Anomalies for the same variables (mean values are average of all data from ~2005, ~2009–2011 or ~2009–2012). The drought period in our drought sites had a slightly lower average temperature during the drought than during the equivalent months of 2009 (24.6 °C versus 24.7 °C). The bar highlights the approximate period of the 2010 drought in the region based on CWD anomaly. To calculate CWD see ref. 40. Error bars are standard error differences between plots.

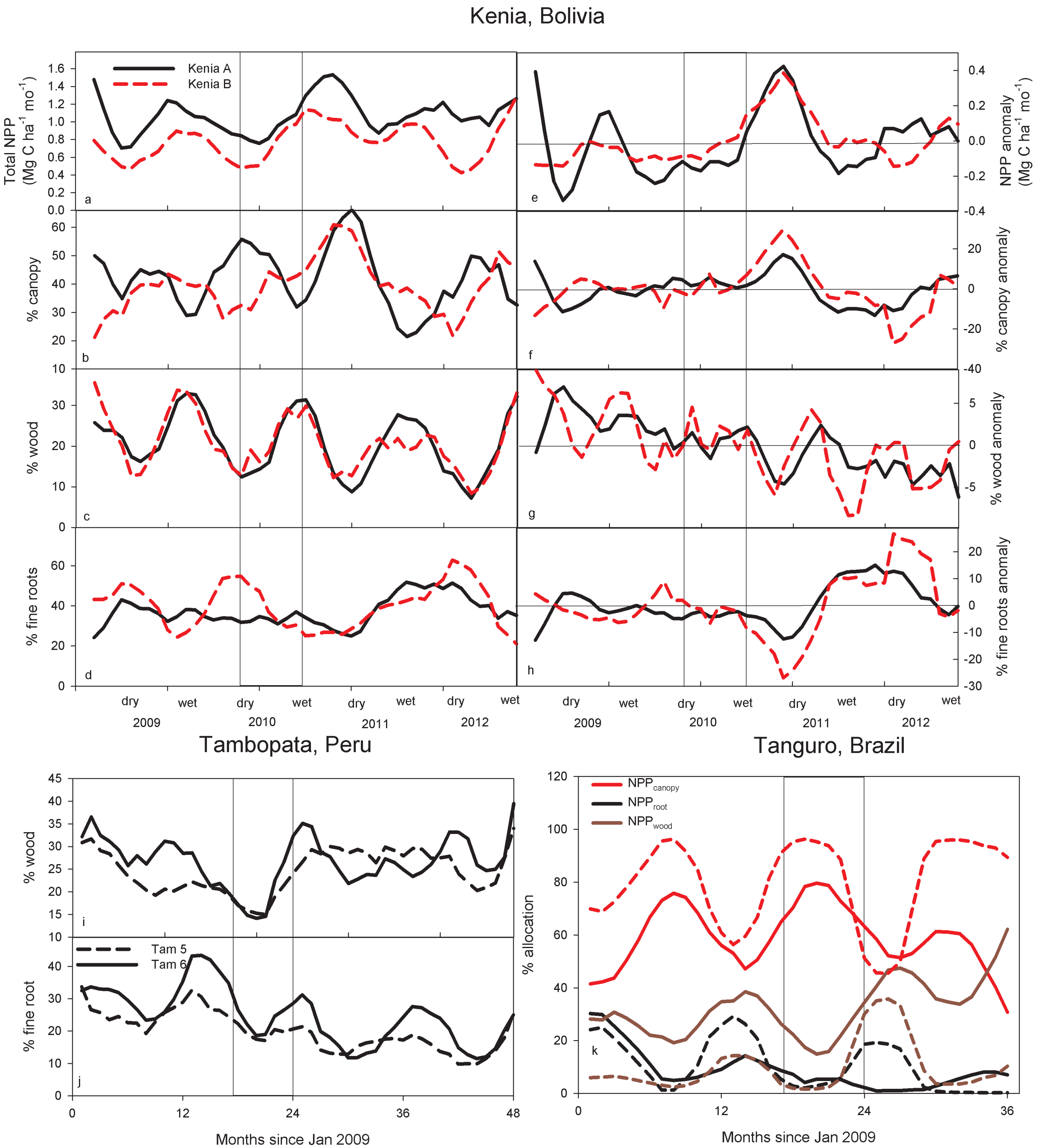
Plot	November 2010 (drought)	June 2011 (no drought)	Change
Kenia-A	2.92 ± 0.85	6.0 ± 0.98*	3.08 μmol m^{-2} s^{-1}
Kenia-B	–0.29 ±0.09	3.1 ± 0.67***	3.39 μmol m^{-2} s^{-1}



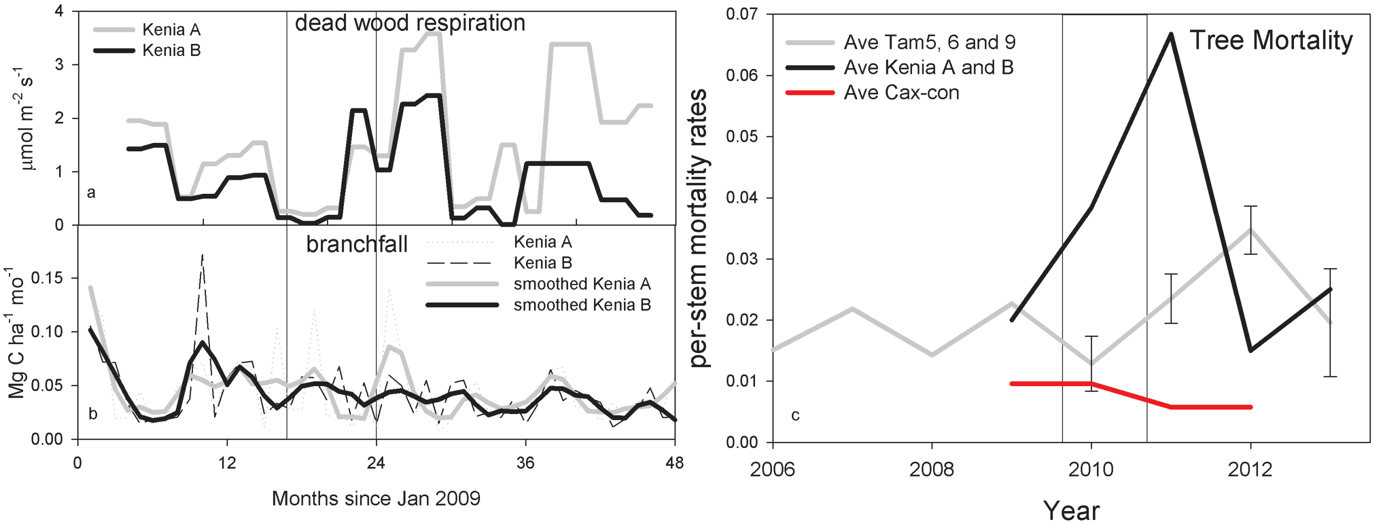
Extended Data Figure 3 | Leaf-level light-saturated photosynthesis measurements. Top: light-saturated (1,000 μmol m^{-2} s^{-1} irradiance, 25 °C, ambient CO_2) leaf gas exchange (μmol m^{-2} s^{-1}) (means ± s.e.m.) for a drought period (November 2010) and a non-drought period (June 2011) for sunlit branches (cut and rehydrated) on the same ~20 trees each season distributed evenly through the two (Kenia-A and Kenia-B) 1-ha plots. Asterisks indicate significant differences between the plots: *$P < 0.05$; ***$P < 0.001$. Bottom: weekly averaged leaf-level photosynthesis for eight species from three canopy walk-up towers measured at 1,000 μmol m^{-2} s^{-1} light and 30 °C between July (the start of the dry season) and November from the Tapajós, Brazil (see ref. 41 for further details and methodology). In the Tapajós the average dry season lasts from about July to about November. Note the lack of a decrease in photosynthesis during the dry season. Over this period, soil moisture decreases from ~0.45 to 0.40 m^3 m^{-3}, most of the decrease that occurs during the dry season[42]. These data suggest that a large (that is, 50%) sharp decline in leaf-level photosynthesis is not typical during an average dry season and that the declines shown in the table above are probably due to the 2010 drought.



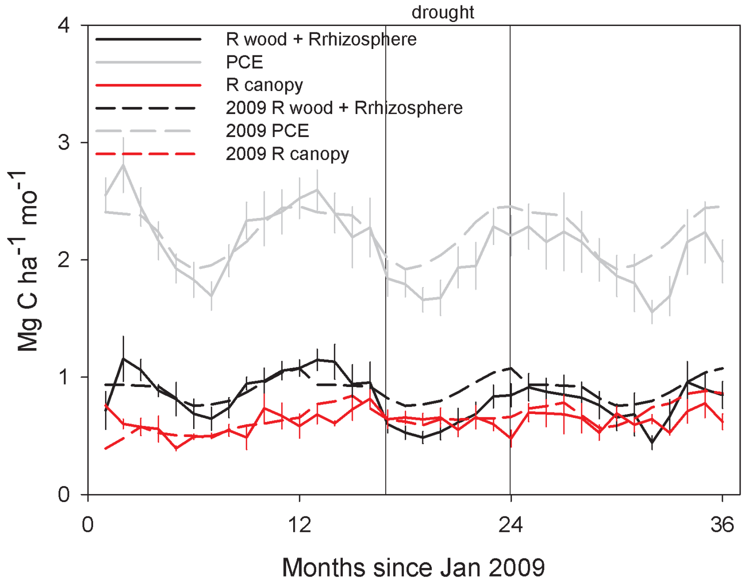
Extended Data Figure 4 | A conceptual model with simulated data of the impact of drought on the study sites. Total photosynthesis (grey dashed line; 100% represents average photosynthesis) decreased during the drought period (vertical bar). Total NPP (grey line, shown as a percentage of total photosynthesis) and growth respiration (R_{growth}) (black dotted line) remained constant, whereas maintenance respiration ($R_{maintain}$) (black line) decreased after NSC stores were depleted. Total NSC stores decreased (we define a negative value as NSC storage) during the drought period (the red line indicates a NSC storage of 0) and then increased at the end of the drought. Red arrows represent the timing of when the basin was a source (up) or a sink (down) of CO_2 to the atmosphere based on atmospheric inversion measurements from ref. 5.



Extended Data Figure 5 | Impact of drought on carbon allocation for individual plots. This figure shows similar trends to those in Fig. 2, but with all the plots separated and extended for a further year for Tambopata and Kenia. **a**–**d**, Four years of total NPP (Mg C ha^{-1} per month) (**a**), percentage allocation to canopy (**b**), percentage allocation to wood (**c**) and percentage allocation to fine roots (**d**) for the two plots in Kenia, Bolivia (black line, Kenia-A; grey line, Kenia-B). **e**–**h**, Seasonally detrended anomaly data for the same variables. This data set is explored in detail in ref. 26. **i**, **j**, Comparison of four years of wood (**i**) and root (**j**) allocation data for two plots in Tambopata, Peru (black lines). Canopy NPP is not shown because LAI data were not processed for the entire 4-year period. **k**, Three years of woody (brown), fine-root (black) and canopy (red) allocation data for two plots in Tanguro, Brazil (solid line, Tanguro A; dashed line, Tanguro C). The bar indicates the approximate drought period.



Extended Data Figure 6 | Deadwood respiration, branch fall and tree mortality. **a**, Respiration from deadwood over a 4-year period from Kenia-A (grey) and Kenia-B (black). **b**, Branch fall over a 4-year period from Kenia-A (grey) and Kenia-B (black); smoothed values are shown in bold lines; actual values are shown in dashed lines. **c**, Per-stem mortality rates for Peruvian drought plots (grey line, $n = 3$; error bars indicate standard errors), Bolivian drought plots (black line, $n = 2$) and the control plot in Caxiuanã (red line, $n = 1$). We do not show mortality for the Brazilian drought plots, but a recent paper[43] has shown an increase in mortality after drought at these sites. Mortality was marginally significantly higher ($P = 0.06$; paired 1-tailed t-test, $n = 5$) during the 2-year period after the drought than in other periods. The bar indicates the approximate period of the drought.



Extended Data Figure 7 | Separated components of PCE. Total PCE (grey solid line), wood and rhizosphere respiration (black solid line), and canopy respiration (red solid line) for the six droughted plots and smoothed 2009 equivalents (stippled lines). This figure shows that the decline in R_a was due to the components measured monthly (wood and rhizosphere respiration) and not to canopy respiration (which was measured only once or twice a year). This does not mean that canopy respiration did not decrease during the drought, only that we did not track canopy respiration sufficiently to measure changes. The bar indicates the approximate period of the drought.

Extended Data Table 1 | Methods for intensive monitoring of net primary production and photosynthesis

	Component	Description	Sampling period	Sampling interval
Net primary productivity (*NPP*)	Above-ground coarse wood net primary productivity (NPP_{ACW})	Forest inventory: All trees ≥ 10 cm DBH censused to determine growth rate of existing surviving trees and rate of recruitment of new trees, and identified floristically to derive wood density. Stem biomass calculated using the Chave et al. (2005) allometric equation for tropical moist forests, employing diameter, height and wood density data[44].	Kenia 2009-2013,Tambopata 2005-2013,Caxiuanã 2009-2010,Wayqecha 2009-2010,San Pedro 2009-2010,Tanguro 2009-2011	Every year (trees ≥ 10 cm DBH)
		Seasonal growth: Dendrometers were installed on most to all trees > 10cmDBH in each plot to determine the spatial-temporal and seasonal variation in growth.	Kenia 2009-2012,Tambopata 2009-2012,Caxiuanã 2009-2010,Wayqecha 2009-2010,San Pedro 2009-2010,Tanguro 2009-2011	Every month to 3 months
	Litterfall net primary productivity ($NPP_{litterfall}$)	Litterfall production of dead organic material less than 2 cm diameter was estimated by collecting litterfall in 0.25 m^2 (50 cm x 50 cm) litter traps placed at 1 m above the ground at the centre of each of the 25 subplots in each plot.	Kenia 2009-2012,Tambopata 2009-2012,Caxiuanã 2009-2010,Wayqecha 2009-2010,San Pedro 2009-2010,Tanguro 2009-2011	Every month
	Leaf area index (LAI)	Canopy images were recorded with a digital camera and hemispherical lens near the centre of each of the 25 subplots in each plot, at a standard height of 1 m during early morning or overcast conditions. LAI was estimated from these images using CAN-EYE software.	Kenia 2009-2012,Tambopata 2009-2012,Caxiuanã 2009-2010, Wayqecha 2009-2010,San Pedro 2009-2010,Tanguro 2009-2011	Every month
	Fine root net primary productivity ($NPP_{fine\ roots}$)	Sixteen ingrowth cores (mesh cages 14 cm diameter, installed to 30 cm depth) were installed in each plot. Cores were extracted and roots were manually removed from the soil samples in four 10 min time steps and the pattern of cumulative extraction over time was used to predict root extraction beyond 40 minutes. Root-free soil was then re-inserted into the ingrowth core. Collected roots were thoroughly rinsed, oven dried at 60°C to constant mass, and weighed.	Kenia 2009-2012,Tambopata 2009-2012,Caxiuanã 2009-2010,Wayqecha 2009-2010,San Pedro 2009-2010,Tanguro 2009-2011	Every 3 months
Leaf Photosynthesis	Saturated leaf photosynthesis (A_{sat})	Photosynthesis at PAR levels of 1000 µmol m^{-2} s^{-1}, 25˚C temperature and ambient CO_2 concentrations (~400ppm) were recorded on ~20 trees per plot with an IRGA and specialized cuvette (Bolivia - Ciras 2, PPsystems, Tapajos – Licor 6400). In Bolivia, during the peak of the drought (November 2011) and again in June 2011, on the same trees during both periods, we randomly selected one branch each from the sunlit portion of the canopy and immediately re-cut the branches underwater to restore hydraulic connectivity. In the Tapajos, measurements were made weekly for 16 weeks on eight, top of canopy tree species accessed using canopy walk up towers (see Doughty 2011 for details)[41].	Kenia Nov 2010 and June 2011, and Tapajos, Brazil starting in June and measured weekly for 16 weeks	variable

See also the RAINFOR-GEM manual 2012 (available at http://gem.tropicalforests.ox.ac.uk/).

Extended Data Table 2 | Methods for intensive monitoring of autotrophic and heterotrophic respiration

	Component	Description	Sampling period	Sampling interval
R_a and R_h	Total soil CO_2 efflux (R_{soil})	Total soil CO_2 efflux was measured using a closed dynamic chamber method, at the centre of each of the 25 subplots in each plot with an infra-red gas analyser (IRGA; EGM-4) and soil respiration chamber (SRC-1) sealed to a permanent collar in the soil.	Kenia 2009-2012,Tambopata 2009-2012,Caxiuanã 2009-2010,Wayqecha 2009-2010,San Pedro 2009-2010,Tanguro 2009-2011	Every month
	Soil CO_2 efflux partitioned into autotrophic ($R_{rhizosphere}$) and heterotrophic ($R_{soilhet}$) components	At four points at each corner of the plot, we placed plastic tubes of 12 cm diameter; three tubes with short collars (10 cm depth) allowing both heterotrophic and rhizosphere respiration, three tubes with longer collars (40 cm depth) with no windows to exclude both roots and mycorrhizae. At the centre of each plot, a control experiment was carried out in order to assess the effects of root severing and soil structure disturbance that occurs during installation.	Kenia 2009-2012,Tambopata 2009-2012,Caxiuanã 2009-2010,Wayqecha 2009-2010,San Pedro 2009-2010,Tanguro 2009-2011	Every month
	Canopy respiration (R_{leaves})	In each plot, leaf dark respiration was recorded for an average of ~20 trees per plot with an IRGA and specialized cuvette. For each tree, we randomly selected one branch each from sunlit and shaded portions of the canopy and immediately re-cut the branches underwater to restore hydraulic connectivity. Leaves were placed in a black bag for at least 10 minutes prior to measurement.	Kenia 2010-2011,Tambopata 2010-2011,Caxiuanã 2007 Wayqecha 2010-2011,San Pedro 2010-2011	Once in dry season, once in wet season
	Above-ground live wood respiration (R_{stems})	Bole respiration was measured using a closed dynamic chamber method, from 25 trees distributed evenly throughout each plot at 1.3 m height with an IRGA (EGM-4) and soil respiration chamber (SRC-1) connected to a permanent collar, sealed to the tree bole surface.	Kenia 2009-2012,Tambopata 2009-2012,Caxiuanã 2009-2010,Wayqecha 2009-2010,San Pedro 2009-2010,Tanguro 2009-2011	Every 1-2 months
	Above-ground dead wood respiration	Representative samples of five decomposition categories of wood were placed in the respiration chamber (SRC-1) of an IRGA (EGM-4). See RAINFOR-GEM manual for description of decomposition status and surface area formulas to scale results to the 1 ha plot.	Kenia 2009-2012,Tambopata 2009-2012,Caxiuanã 2009-2010,Wayqecha 2009-2010,San Pedro 2009-2010,Tanguro 2009-2011	Every month to two months

See also the RAINFOR-GEM manual 2012 (available at http://gem.tropicalforests.ox.ac.uk/).

Extended Data Table 3 | Data analysis techniques for intensive study of carbon dynamics

	Component	**Data processing details**
Net primary productivity (NPP_{AG})	Above-ground coarse wood net primary productivity (NPP_{ACW})	Biomass calculated using the Chave *et al.* (2005)[39] allometric equation for tropical moist forests: $AGB = 0.0509 \times (\rho D^2 H)$ where AGB is aboveground biomass (kg), ρ is density (g cm^{-3}) of wood, D is dbh (cm), and H is (measured or if missing based on Feldpausch et al. 2012[36]) height (m). To convert biomass values into carbon, we assumed that dry stem biomass is 48-50% carbon based on the following reference[45].
	Branch turnover net primary productivity ($NPP_{branch\ turnover}$)	See RAINFOR-GEM manual for description of decomposition status and surface area formulas.
	Litterfall net primary productivity ($NPP_{litterfall}$)	Litterfall is separated into different components, oven dried at 60°C to constant mass and weighed. Litter is estimated to be 49.2% carbon, based on mean Amazonian values (S. Patiño, unpublished analysis).
	Leaf area index (LAI)	LAI was estimated using "true LAI" output from the Caneye program which accounts for clumping of foliage, and assumes a fixed leaf inclination angle, based on average estimates at individual plots. Leaves were separated into sunlit and shaded fractions using the following equation: $F_{sunlit} = (1-\exp(-K*LAI))/K$ where K is the light extinction coefficient, and F_{sunlit} is the sunlit leaf fraction[31]. The model assumptions are randomly distributed leaves, and $K = 0.5/\cos(Z)$ where Z is the solar zenith angle, which was set to 30° in this study.
	Fine root net primary productivity ($NPP_{fine\ roots}$)	Roots were manually removed from the soil samples in four 10 min time steps, according to a method that corrects for underestimation of biomass of hard-to-extract roots [46] and used to predict root extraction beyond 40 min (up to 100 minutes); we estimate that there was an additional 28% correction factor for fine roots not collected within 40 minutes. Correction for fine root productivity below 30 cm depth increased the value by 39%.
Autotrophic and heterotrophic respiration	Total soil CO_2 efflux (R_{soil})	Soil surface temperature (T260 probe, Testo Ltd, Hampshire, U.K.) and moisture (Hydrosense probe, Campbell Scientific Ltd, Loughborough, UK) were recorded at each point after efflux measurement.
	Soil CO_2 efflux partitioned into autotrophic ($R_{rhizosphere}$) and heterotrophic ($R_{soilhet}$) components	The partitioning experiment allows estimation of the relative contributions of (1) surface organic litter, (2) roots, (3) mycorrhizae and (4) soil organic matter to total soil CO_2 efflux. Contributions are estimated from differences between collars subjected to different treatments, in excess of pre-existing spatial variation.
	Canopy respiration (R_{leaves})	To scale to whole-canopy respiration, mean dark respiration for shade and sunlit leaves were multiplied by the respective estimated fractions of total LAI. To account for daytime light inhibition of leaf dark respiration, we apply the inhibition factor applied in Malhi *et al.* (2009) [47] (67% of daytime leaf dark respiration, 33% of total leaf dark respiration).
	Above-ground live wood respiration (R_{stems})	To estimate plot-level stem respiration tree respiration per unit bole area was multiplied by bole surface area (SA) for each tree, estimated with the following equation[48]: $\log(SA) = -0.105 - 0.686 \log(DBH) + 2.208 \log(DBH)^2 - 0.627 \log(DBH)^3$ Where DBH is bole diameter at 1.3 m height.

See also the RAINFOR-GEM manual 2012 (available at http://gem.tropicalforests.ox.ac.uk/).

doi:10.1038/nature14224

Reconstructed *Homo habilis* type OH 7 suggests deep-rooted species diversity in early *Homo*

Fred Spoor[1,2]*, Philipp Gunz[1]*, Simon Neubauer[1], Stefanie Stelzer[1], Nadia Scott[1], Amandus Kwekason[3] & M. Christopher Dean[2]

Besides *Homo erectus* (*sensu lato*), the eastern African fossil record of early *Homo* has been interpreted as representing either a single variable species, *Homo habilis*[1], or two species[2–6]. In the latter case, however, there is no consensus over the respective groupings, and which of the two includes OH 7, the 1.8-million-year-old *H. habilis* holotype[7]. This partial skull and hand from Olduvai Gorge remains pivotal to evaluating the early evolution of the *Homo* lineage, and by priority names one or other of the two taxa. However, the distorted preservation of the diagnostically important OH 7 mandible has hindered attempts to compare this specimen with other fossils[8,9]. Here we present a virtual reconstruction of the OH 7 mandible, and compare it to other early *Homo* fossils. The reconstructed mandible is remarkably primitive, with a long and narrow dental arcade more similar to *Australopithecus afarensis* than to the derived parabolic arcades of *Homo sapiens* or *H. erectus*. We find that this shape variability is not consistent with a single species of early *Homo*. Importantly, the jaw morphology of OH 7 is incompatible with fossils assigned to *Homo rudolfensis*[8] and with the A.L. 666-1 *Homo* maxilla. The latter is morphologically more derived than OH 7 but 500,000 years older[10], suggesting that the *H. habilis* lineage originated before 2.3 million years ago, thus marking deep-rooted species diversity in the genus *Homo*. We also reconstructed the parietal bones of OH 7 and estimated its endocranial volume. At between 729 and 824 ml it is larger than any previously published value, and emphasizes the near-complete overlap in brain size among species of early *Homo*. Our results clarify the *H. habilis* hypodigm, but raise questions about its phylogenetic relationships. Differences between species of early *Homo* appear to be characterized more by gnathic diversity than by differences in brain size, which was highly variable within all taxa.

Computed tomography (CT) reveals that distortion of the OH 7 mandible mostly involves a small number of fractures and displacements, without evidence of larger-scale plastic deformation (Fig. 1a). In a CT-based virtual reconstruction the individual parts of the corpus were therefore realigned by simple repositioning. The dental arcade from the left M_2 to the right M_1 was restored using combined clues from the interstitial facets, the alveoli and the observation that the displacement of the incisors occurred along a single transverse crack through their roots (Fig. 1d and Extended Data Figs 1–3). The dislocated right corpus covers a well-preserved lingual symphyseal surface with a marked genial spine and bilateral genioglossal pits (Fig. 1c). These features and the symmetrically placed apical halves of the I_1 roots jointly mark the midsagittal plane (Fig. 1b). Using this plane a second reconstruction was made by mirror-imaging the left side from M_2 to C to the right (Fig. 1f). The CT scans show that root formation of the C and M_2 was incomplete (Fig. 1c; contrary to ref. 9), confirming the original suggestion that OH 7 was a late juvenile[11], rather than an adult[12] (Supplementary Note 1). Conclusions formed on the basis of the specific morphological analyses presented below are unlikely to be affected by the sub-adult status of OH 7 and some other early *Homo* fossils (Extended Data Figs 4 and 5 and Supplementary Note 2).

The two mandibular reconstructions of OH 7 are characterized by relatively long and parallel post-canine tooth rows. This primitive morphology contrasts with several other early *Homo* specimens, which show derived parabolic arcade shapes (Extended Data Fig. 6). We therefore test whether this shape variability is consistent with a single species of early *Homo* (*H. erectus*, following ref. 13). Alternative evolutionary scenarios invoke multiple species of early *Homo* that differ, at least, in dental arcade shape. We used geometric morphometrics to quantify the shape variability of dental arcade shape in early *Homo*, in the context of (1) the

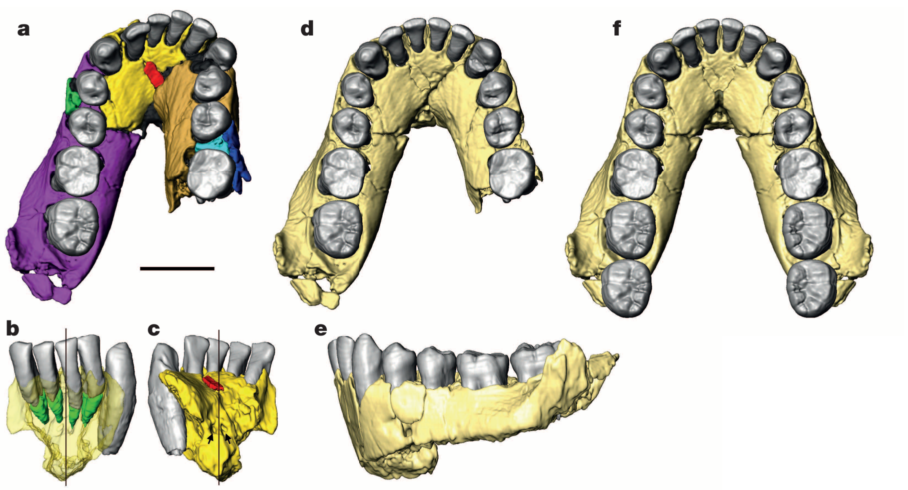


Figure 1 | CT-based visualization of the OH 7 mandible. **a**, As preserved, marking individual parts that were adjusted in the reconstruction. **b**, **c**, The anterior corpus showing the midsagittal plane in relation to the incisor roots (**b**, transparent, anterior view), and the genial spine and pits (**c**, posterior view; arrows indicate pits). **d**, **e**, The reconstructed mandible in occlusal and left lateral views, respectively. **f**, The mandible reconstructed using the mirror-imaged left corpus; M_2s duplicated as M_3s, and a small damaged area in the midline of the post-incisive plane shaded. Scale bar is 2 cm.

[1]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [2]Department of Cell and Developmental Biology, University College London, London WC1E 6BT, UK. [3]Museum and House of Culture, National Museum of Tanzania, Dar es Salaam, Tanzania.
*These authors contributed equally to this work.

within-group shape variability of extant apes and humans, and (2) the effect of allometry on dental arcade shape. These analyses show that the shape of the OH 7 arcade is close to that of the early *Homo* mandibles KNM-ER 1802 and OH 13, and is more similar to shapes found in great apes and *Australopithecus* than in *H. sapiens* (Fig. 2a and Extended Data Figs 4 and 5). Compared with other early *Homo* specimens it strongly contrasts with mandibles of *H. erectus* and with KNM-ER 1482 and KNM-ER 60000, specimens that have been associated with the iconic cranium KNM-ER 1470 (ref. 8; Fig. 2a, b). This difference is as large as between *Gorilla gorilla* and *Pan troglodytes.* When specimens are attributed to *H. habilis* (OH7, OH 13, KNM-ER 1802), *H. erectus* (Dmanisi) and *H. rudolfensis* (KNM-ER 1482, KNM-ER 60000; following ref. 8) the respective within-group shape variations are consistent with those shown by extant hominids (Fig. 2c). In contrast, pooling any of these early *Homo* groups yields shape distances that far exceed those found in extant species, and distributions are mostly bimodal (Fig. 2c). We found no static allometric effects on dental arcade shape (Extended Data Fig. 7).

To compare OH 7 with key maxillae attributed to early *Homo* we predicted the shape of its maxillary dental arcade using multiple multivariate regression analysis, on the basis of the strong and consistent pattern of covariation between the maxillary and mandibular dental arcade shapes of extant humans and great apes (Extended Data Figs 8 and 9). In addition, we manually reconstructed the maxillary arcade of OH 7 on the basis of an occlusal match between lower and upper tooth crowns (Extended Data Fig. 6c). The predicted maxillary arcade follows the same morphological pattern as the mandible, is most similar in shape to that of *A. afarensis*, and is different from those of other early *Homo* maxillae (Fig. 2d, e and Extended Data Figs 4 and 5). The inferred OH 7 maxilla is most similar to KNM-ER 1813, but at the extreme boundaries of the intraspecific shape differences found between two individuals in extant humans and great apes (Fig. 2e). The difference between the *H. erectus* maxilla D2282 and OH 7 falls just within the 95% boundary of *G. gorilla*, but the better preserved mandible D211 of the same individual clearly differs from OH 7 (Fig. 2c). OH 7 is most distinct from KNM-ER 62000, a specimen that closely resembles KNM-ER 1470 (ref. 8). This shape difference is as large as that between some extant *H. sapiens* and *Pan troglodytes* individuals (Fig. 2e) and variation of the pooled early *Homo* sample well exceeds that of extant hominid taxa (Extended Data Fig. 5a).

The left and right parietal bones of OH 7 are both incomplete, but bilaterally preserved parts complement each other[1], enabling reconstruction of the entire parietal portion of the vault (Fig. 3a, b and Extended Data Fig. 10). In extant humans and great apes the endocranial volume (ECV) is strongly correlated with the form (combined shape and size) of the parietal endocranial surface. Using this relationship and two methods we estimated four separate ECVs from the reconstructed OH 7 parietals, with values between 729 and 824 ml (combined uncertainty range 630–967 ml). These ECVs are considerably larger than previous estimates[1,14–20], typically given as 647–687 ml[1,20] (Fig. 3c). We consider our values more reliable as they are calculated on the basis of larger and taxonomically

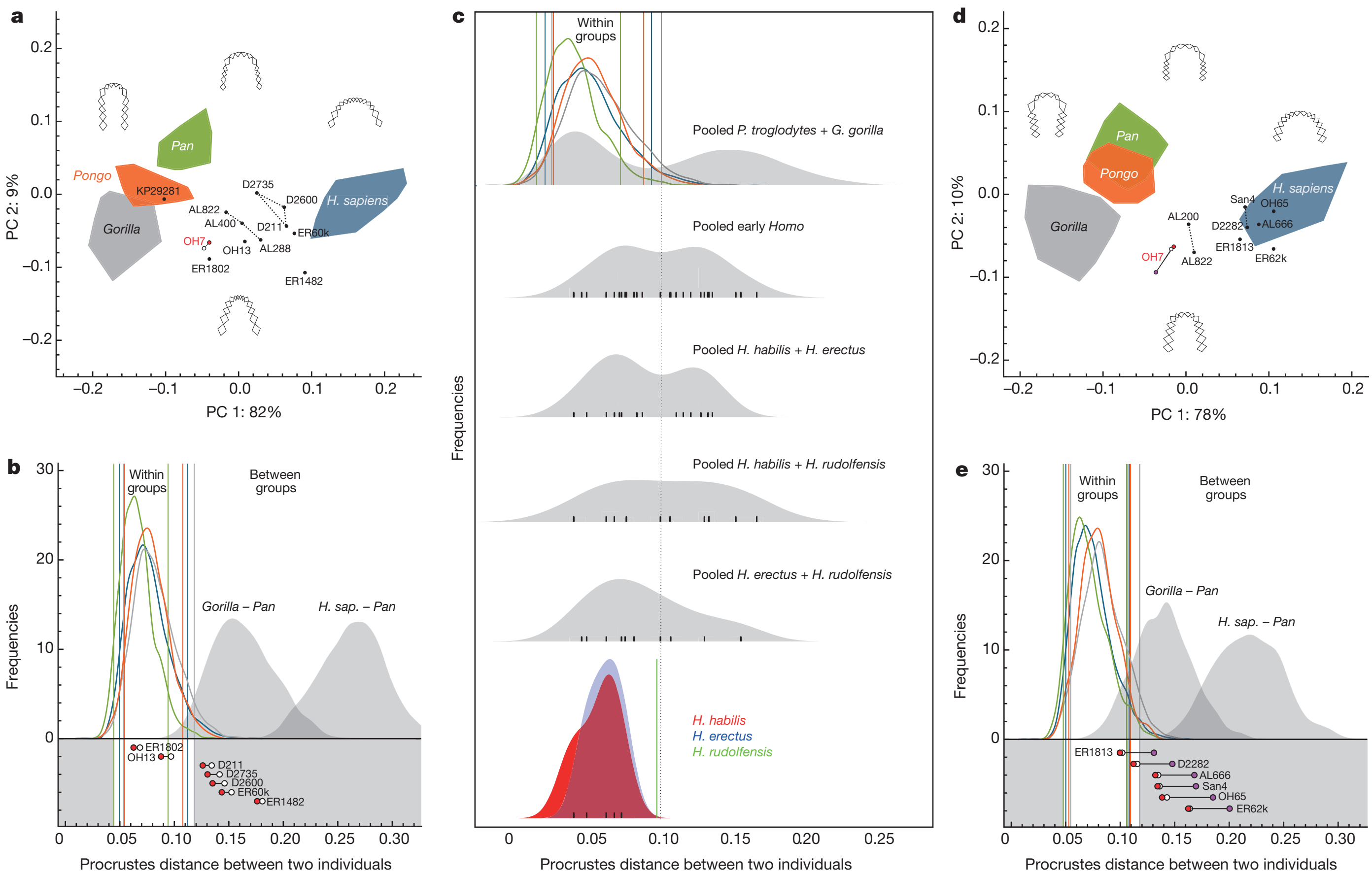


Figure 2 | Analysis of the OH 7 jaw. a, Plot of principal components (PC) 1 and 2 describing the shape of the mandibular dental arcade; associated shape changes three standard deviations away from the mean are shown. OH 7 is represented by the full-corpus (red) and symmetry-based (white) reconstructions. **b**, Procrustes distances between OH 7 and other early *Homo* mandibles compared to frequency plots of distances between all possible individual pairs within and between extant groups (colours as in **a**; lines represent 5% and 95% limits of intraspecific distributions). **c**, Frequency plots of the mandibular shape differences between all pairs within extant groups (as in **b**), a pooled sample of chimps and gorillas (top), and various groupings of early *Homo* fossils (lower five), with black ticks showing the distance between each pair. **d**, Same as **a** for maxillary shape. For OH 7 the two statistical predictions and one anatomical prediction (purple) are plotted. **e**, Same as **c** for pairwise shape differences between maxillae. As in **d** three reconstructions of OH 7 are plotted.

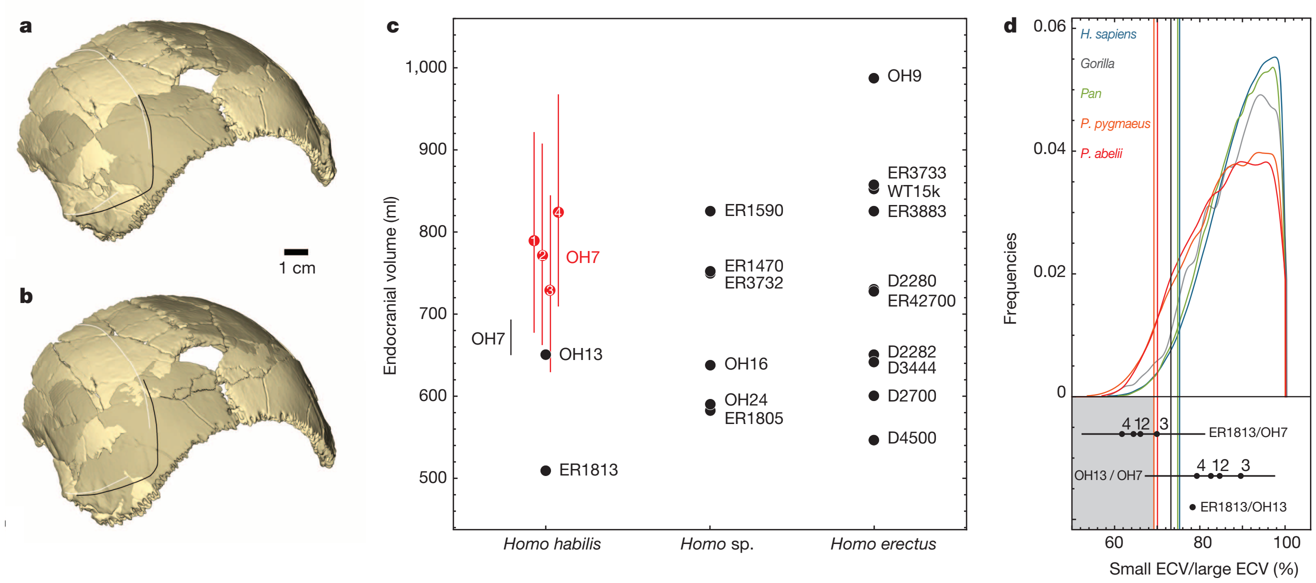


Figure 3 | Reconstruction and analysis of the OH 7 parietals. a, **b**, Two alternative anatomical reconstructions in posterolateral view, with temporal lines shown in white (left side) and black (mirrored right side), mirror-imaged bones in darker shade. **c**, Means and ranges of four ECV predictions for OH 7 (red 1 and 2, regression-based and geometric reconstruction based estimates for first anatomical reconstruction; red 3 and 4, corresponding estimates for second anatomical reconstruction), compared with typical values reported previously[1,19] (black), and with other early *Homo* specimens. **d**, Difference in ECV between OH 7, OH 13 and KNM-ER 1813 compared with pairwise intraspecific differences within extant hominid species as a ratio of the smaller ECV divided by the larger ECV. Lines represent the 5% limits of these distributions.

more diverse comparative samples, and the linear regression model is not taxon-dependent, performing equally well for great ape species and humans (Extended Data Fig. 10f).

The ECV estimates obtained for OH 7 fall well within the range of early *H. erectus* (546–1067 ml) and are among the largest values reported for other specimen attributed to early *Homo*[13,20] (Fig. 3c). Specimens that show affinities with OH 7 in dental arcade shape and have estimated ECVs are OH 13 and KNM-ER 1813. The difference in ECV between OH 13 (650 ml, ref. 15) and OH 7 falls well within the intraspecific variation of extant hominids (Fig. 3c). The difference between KNM-ER 1813 (509 ml, ref. 20) and OH 7 is substantially larger, and can only be sampled in extant humans and great apes when the lower ECV predictions are considered for OH 7.

The reconstructed mandible and parietals of OH 7 reveal that the holotype of *H. habilis* is characterized by primitive gnathic morphology; a long and narrow dental arcade indicative of a subnasally prognathic face, combined with a larger ECV than assumed thus far. Analysing shape variation of the dental arcades shows that the early *Homo* sample represents more than one species (contrary to ref. 13), and that a *H. habilis* hypodigm of OH 7 combined with KNM-ER 1802 and OH 13 is consistent with intraspecific variation seen in extant hominids (Fig. 2c). Both gnathic and parietal evidence indicate that it is possible that KNM-ER 1813 is also part of this hypodigm. If so, *H. habilis* would have been variable, suggesting considerable sexual dimorphism as well as evolutionary changes over the documented temporal range of these specimens from 2 million years (Myr) to 1.65 Myr ago[21,22].

In terms of arcade shape OH 7 is most distinct from the partial cranium KNM-ER 62000 and the mandible KNM-ER 1482. These specimens have both been associated with KNM-ER 1470, and are marked by shorter post-canine tooth rows and a non-projecting anterior row[9]. The KNM-ER 1470 maxilla is incompletely preserved and could not be included in the analyses, but visual comparison with OH 7 nevertheless suggests a comparable difference in arcade shape (Extended Data Fig. 6h, i). This major difference in gnathic morphology fully supports the view that KNM-ER 1470 and associated specimens are best attributed to a species other than *H. habilis* (contrary to ref. 13) for which the name *H. rudolfensis* is available[23]. The mandible KNM-ER 60000 has been associated with KNM-ER 1470, KNM-ER 62000 and KNM-ER 1482 (ref. 8), but the shape of its dental arcade shows the closest similarity to early *H. erectus* mandibles from Dmanisi rather than to KNM-ER 1482 (Fig. 2a). Clearly, the relationship between *H. rudolfensis* and early *H. erectus* warrants further quantitative analysis using an expanded landmark data set which comprehensively captures the facial features reported to characterize *H. rudolfensis*[3,4,8] but this lies beyond the scope and focus of this paper.

The maxillae OH 65 and A.L. 666-1 have been associated with *H. habilis*[6,10]. However, considered by the standards of intraspecific variation among extant hominids, their morphology is incompatible with OH 7 (Fig. 2e), and hence *H. habilis*. Both are close to modern humans in dental arcade shape (Fig. 2d and Extended Data Fig. 4c, d), but lack the unique facial morphology used to diagnose *H. rudolfensis*[8,10]. Suggested similarities to early *H. erectus*[13] are consistent with the evidence presented here, but require further examination. Regardless of species affinities it is remarkable that the 2.3-Myr-old[10] A.L. 666-1 is 0.5 Myr older than OH 7, but is distinctly more derived in gnathic morphology. This pattern suggests that the *H. habilis* lineage originated before 2.3 Myr ago, predating the oldest fossil evidence for the genus *Homo*[10]. OH 13 shows that *H. habilis* survived until at least 1.65 Myr ago, but a previously reported last appearance date of 1.44 Myr[21] can no longer be demonstrated in light of new evidence presented here and elsewhere[13] (Supplementary Note 3).

Our new ECV estimates of OH 7 support the view that for the period between 2.1 and 1.5 Myr ago the ECVs of *H. habilis*, *H. rudolfensis* and *H. erectus* largely overlapped in range[24], and broadly fell between 500 and 900 ml (Fig. 3c). Hence, although *H. erectus* shows autapomorphic features of the calvaria[21], there is little evidence that the early radiation of the genus *Homo* was marked by interspecific diversification in brain size. In contrast, it is facial morphology that appears to distinguish these three species. The dental arcades of *H. habilis* show the most primitive morphology, with long and parallel post-canine dental rows suggestive of a subnasally prognathic face, as opposed to the more diverging and reduced post-canine rows seen in early *H. erectus*, *H. sapiens* and A.L. 666-1, the currently earliest known *Homo*. The dental arcade of the *H. rudolfensis* face KNM-ER 62000 also shows shorter post-canine rows, but these are more parallel, and the incisor row is distinctly flat and retracted along the bi-canine line (Extended Data Fig. 5b and

Supplementary Note 2). Brain size increase in the genus *Homo*, and *H. erectus* in particular, has been linked with a shift towards energy- and nutrient-rich diets and a reduced masticatory system[25–28]. Nevertheless, the *H. habilis* type specimen, OH 7, demonstrates that increased brain size, equivalent to that in early *H. erectus*, occurred in combination with a more primitive gnathic morphology without any signs of reduced dental size or corpus robusticity (Supplementary Note 4).

Note added in proof: The recently described LD 350-1 partial mandible from Ethiopia now provides the earliest evidence of the genus *Homo* at 2.8 Myr (Villmoare, B. *et al.* Early *Homo* at 2.8 Ma from Ledi-Geraru, Afar, Ethiopia. *Science*, in the press). Its morphology is more primitive than seen in the mandibles we attribute to *H. habilis*, and this new evidence supports the conclusions of our study.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 April 2014; accepted 8 January 2015.

1. Tobias, P. V. *Olduvai Gorge Volume 4: The Skulls and Endocasts of* Homo habilis (Cambridge Univ. Press, 1991).
2. Stringer, C. B. in *Major Topics in Primate and Human Evolution* (eds Wood, B., Martin, L. & Andrews, P.) 266–294 (Cambridge Univ. Press, 1986).
3. Lieberman, D. E., Pilbeam, D. R. & Wood, B. A. A probabilistic approach to the problem of sexual dimorphism in *Homo habilis*: a comparison of KNM-ER 1470 and KNM-ER 1813. *J. Hum. Evol.* **17,** 503–511 (1988).
4. Wood, B. Origin and evolution of the genus *Homo. Nature* **355,** 783–790 (1992).
5. Rightmire, G. P. Variation among early *Homo* crania from Olduvai Gorge and the Koobi Fora region. *Am. J. Phys. Anthropol.* **90,** 1–33 (1993).
6. Blumenschine, R. J. *et al.* Late Pliocene *Homo* and hominid land use from western Olduvai Gorge, Tanzania. *Science* **299,** 1217–1221 (2003).
7. Leakey, L. S. B., Tobias, P. V. & Napier, J. R. A new species of the genus *Homo* from Olduvai Gorge. *Nature* **202,** 7–9 (1964).
8. Leakey, M. G. *et al.* New fossils from Koobi Fora, northern Kenya, confirm taxonomic diversity in early *Homo. Nature* **488,** 201–204 (2012).
9. Antón, S. C., Potts, R. & Aiello, L. C. Evolution of early *Homo*: an integrated biological perspective. *Science* **345** (2014).
10. Kimbel, W. H., Johanson, D. C. & Rak, Y. Systematic assessment of a maxilla of *Homo* from Hadar, Ethiopia. *Am. J. Phys. Anthropol.* **103,** 235–262 (1997).
11. Leakey, L. S. B. New finds at Olduvai Gorge. *Nature* **189,** 649–650 (1961).
12. Schwartz, J. H. & Tattersall, I. *The human fossil record*, Vol. 2 (Wiley-Liss, 2003).
13. Lordkipanidze, D. *et al.* A complete skull from Dmanisi, Georgia, and the evolutionary biology of early *Homo. Science* **342,** 326–331 (2013).
14. Tobias, P. V. Cranial capacity in anthropoid apes, *Australopithecus* and *Homo habilis*, with comments on skewed samples. *S. Afr. J. Sci.* **64,** 81–91 (1968).
15. Holloway, R. L. New endocranial values for the East African early hominids. *Nature* **243,** 97–99 (1973).
16. Holloway, R. L. The OH 7 (Olduvai Gorge, Tanzania) hominid partial brain endocast revisited. *Am. J. Phys. Anthropol.* **53,** 267–274 (1980).
17. Wolpoff, M. H. Cranial capacity estimates for Olduvai Hominid 7. *Am. J. Phys. Anthropol.* **56,** 297–304 (1981).
18. Holloway, R. L. The OH 7 (Olduvai Gorge, Tanzania) parietal fragments and their reconstruction: a reply to Wolpoff. *Am. J. Phys. Anthropol.* **60,** 505–516 (1983).
19. Vaišnys, J. R., Lieberman, D. & Pilbeam, D. An alternative method of estimating the cranial capacity of Olduvai Hominid 7. *Am. J. Phys. Anthropol.* **65,** 71–81 (1984).
20. Holloway, R. L., Broadfield, D. C. & Yuan, M. S. *The human fossil record*, Vol. 3 (Wiley-Liss. 2004).
21. Spoor, F. *et al.* Implications of new early *Homo* fossils from Ileret, east of Lake Turkana, Kenya. *Nature* **448,** 688–691 (2007).
22. Joordens, J. C. A. *et al.* Improved age control on early *Homo* fossils from the upper Burgi Member at Koobi Fora, Kenya. *J. Hum. Evol.* **65,** 731–745 (2013).
23. Wood, B. '*Homo rudolfensis*' Alexeev, 1986 – fact or phantom? *J. Hum. Evol.* **36,** 115–118 (1999).
24. Lieberman, D. *Evolution of the human head* (Harvard Univ. Press, 2011).
25. Wrangham, R. W., Holland Jones, J., Laden, G., Pilbeam, D. & Conklin-Brittain, N. The raw and the stolen. Cooking and the ecology of human origins *Curr. Anthropol.* **40,** 567–594 (1999).
26. Stedman, H. H. *et al.* Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature* **428,** 415–418 (2004).
27. Leonard, W. R., Snodgrass, J. J. & Robertson, M. L. Effects of brain evolution on human nutrition and metabolism. *Annu. Rev. Nutr.* **27,** 311–327 (2007).
28. Jiménez-Arenas, J. M., Pérez-Claros, J. A., Aledo, J. C. & Palmqvist, P. On the relationships of postcanine tooth size with dietary quality and brain volume in primates: implications for hominin evolution. *BioMed Res. Int.* **2014,** 1–11 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the National Museum of Tanzania, the Tanzania Commission for Science and Technology and the National Museums of Kenya for giving access to fossils in their care, and the Imaging Plus Medical Centre, Dar es Salaam, for CT scanning facilities. We are grateful to M. Leakey, L. Leakey, J.-J. Hublin and S. Antón for support and encouragement, and to R. Blumenschine, P. Corujo, R. David, P. Gokarn, W. Kimbel, K. Kupczik, R. Leakey, J. Lewis, E. Mbua, R. McCarthy, M. Meyer, P. Mitteroecker, P. Msemwa, J. Njau, D. Reinhardt, L. Schroeder, M. Skinner, A. Stoessel, A. Strauss, H. Temming and B. Wood for help with aspects of this study. Research was supported by the Max Planck Society.

Author Contributions F.S., S.N., S.S., N.S., A.K. and C.D. collected data. F.S., P.G., S.N. and C.D. performed analyses. F.S. wrote the paper, with contributions from P.G., S.N. and C.D.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.S. (f.spoor@ucl.ac.uk) or P.G. (gunz@eva.mpg.de).

METHODS

CT scanning. In December 2010 OH 7 was CT scanned with a Siemens Somatom 4, using a slice collimation of 1.0 mm. The image stacks used for 3D visualizations were reconstructed with a voxel size of $0.2 \times 0.2 \times 0.5$ mm, H70h kernel and extended CT scale. The reconstructions of the mandible and parietals use images in the transverse and coronal planes, respectively. OH 7X (right M_2; ref. 29) is housed in the National Natural History Museum in Arusha rather the Museum and House of Culture in Dar es Salaam, and was studied but not CT scanned. Specimens in the comparative sample were scanned with a medical CT scanner (parameters similar to those for OH 7) or with a BIR ACTIS 225/300 (Max Planck Institute for Evolutionary Anthropology, Leipzig), isotropic voxel size of 0.1 mm or less. Avizo 6.3-7.1 (Visualization Sciences Group) and Geomagic Studio 2013 (Geomagic Inc.) were used for visualization, segmentation, reconstruction and 3D landmarking.

Reconstruction of the OH 7 mandible. The reconstruction procedure is discussed in more detail in the Supplementary Methods. The OH 7 mandible was examined by F.S. and M.C.D. to map the pattern of externally visible cracks, displacements and crucial morphological details such as interstitial facets of the tooth crowns. This information was combined with the internal morphology as shown in CT images. The main distortion concerns displacement of teeth and corpus parts along three well-defined fractures: (1) labiolingually through the right C alveolus; (2) through the left corpus from P_4 lingually to C labially; and (3) transversely through the anterior corpus and tooth roots, marking the dislocation of the I crowns (Extended Data Figs 1 and 2).

The 13 teeth and seven corpus parts were digitally separated, as colour coded in Extended Data Fig. 1. The fragmented buccal wall on the right and remnants of matrix were removed, and six corpus parts realigned, guided by their edges and surfaces. Subsequently, the incisors could be realigned, on the basis of the interstitial facets and the apical portion of their roots that remain in their original alveolar positions. The other teeth were repositioned on the basis of their interstitial facets and alveoli. Only the mesial half of the right corpus is preserved, and a second, more complete reconstruction was therefore made, using a copy of the left corpus instead. The latter was mirror-imaged across the midsagittal plane of the anterior corpus, marked by the genial spine and genioglossal pits on the posterior symphyseal surface in combination with the apical halves of the I_1 roots (Fig. 1b, c). The right corpus surrogate thus obtained and the actual right corpus fragment are close in orientation, with the latter being 1 degree more divergent relative to the midsagittal plane. The good match between individual parts confirms that the corpus experienced little or no plastic deformation and that the estimated midsagittal plane is a good reflection of bilateral symmetry (Extended Data Fig. 2e). Finally, to facilitate visual comparisons of overall dental arcade shape the left M_2 was used to represent the similarly sized right M_2, and the missing M_3s (Fig. 1f). However, the M_3s were not considered in any of the shape analyses of the dental arcade. Measurements of the mandibular reconstructions are given in Supplementary Note 5.

Shape analyses of the mandibular and maxillary dental arcades. *Landmarks*. Dental arcades were quantified using 3D landmarks, taken either directly from original specimens or casts with a Microscribe digitizer (Solution Technologies), or from digital volume or surface data using Avizo 6.3 or 7.1 (Visualization Sciences Group). Eighty landmarks are defined which describe the mesial, distal, lingual and buccal/labial alveolar margins of the maxillary and mandibular dental arcade from left to right M3 (Extended Data Fig. 2f, g). The mesial and distal roots of the molars and premolars were landmarked separately, but in the analyses their coordinates were averaged. Mesial and distal landmarks between the tooth crowns are difficult to reach with the Microscribe. These are therefore calculated from pairs of landmarks positioned slightly more buccally/labially and lingually. The M_3s of OH 7 are not preserved and analyses therefore use a subset of landmarks from left to right M2. Results of comparative analyses with or without M3s are not meaningfully different. For visualization the 43 M2-to-M2 landmarks are connected by a wireframe. *Extant species*. Specimens are pairs of matching maxillae and mandibles, without distinct pathologies or malocclusion. Late juveniles are specimens similar in developmental age to OH 7, with erupted Cs and unerupted M3 s. The number of late juveniles that could be included was low because among modern *Homo sapiens* this age group is sparsely represented in museum collections, and among great apes this pattern of dental eruption is rare because eruption of the M3 usually precedes or coincides with eruption of the canine. To maximize sample sizes the late juveniles therefore include a few crania without mandibles.

Homo sapiens: adults 80, worldwide populations; late juveniles 7 maxillae, 5 mandibles. *Pan troglodytes*: adults 41, including *P. t. troglodytes*, *P. t. verus* and *P.t. schweinfurthii*; late juveniles 11 maxillae, 10 mandibles. *Gorilla gorilla*: adults 46; late juveniles 4 maxillae, 2 mandibles. *Pongo* sp.: adults 22.

Fossil specimens. Sample composition of early *Homo* reflects preservation of the dental arcade, a developmental age marked by erupted Cs, and availability of data. *Australopithecus* specimens were included to represent the ancestral or primitive state of gnathic morphology relative to the genus *Homo*. Specimens are labelled in the graphs by their accession number, leaving off the prefix 'KNM-' for specimens housed in the National Museums of Kenya, and the designation '-1' for A.L. specimens housed in the National Museum of Ethiopia.

Australopithecus anamensis KNM-KP29281: CT scan. Landmarks taken from the right side and mirror imaged using the midsagittal plane of the symphysis.

Australopithecus afarensis: A.L. 200-1, A.L. 288-1 and A.L. 400-1: CT-based 3D surface of plaster casts (source: Cleveland Museum of Natural History). Missing landmark data for the left M_2 of A.L. 288-1 estimated using reflected relabelling[30]. A.L. 822-1: casts of the reconstructed cranium and mandible[31], courtesy of W. Kimbel and Y. Rak.

Homo erectus: D211, D2282, D2600, D 2735: CT-based 3D surface of casts (source: Georgian National Museum). The right side of D2282 landmarked and mirror-imaged using the midsagittal plane of the palate. Missing landmarks of the left P_4 and M_1 alveoli of D2600 mirror-imaged from the right and matched into the left tooth row on the basis of the shared P_3 and M_3 landmarks. Sangiran 4: CT scan. Landmarks taken from the right side and mirror-imaged using the midsagittal plane of the palate.

Homo habilis: OH 7: CT-based 3D reconstructions (this paper). Missing landmark data for the M_1 (buccal, distal) and M_2 (all) of the partial right corpus were estimated using reflected relabelling[30]. The second reconstruction employing mirror-imaging provides all landmarks.

Homo sp.: A.L. 666-1: orthogonal, parallel projected 3D surface views of a cast[20]. KNM-ER 1482: CT scan. Landmarks taken from the left side and mirror-imaged using the midsagittal plane of the symphysis. KNM-ER 1802, KNM-ER 60000, KNM-ER 62000: CT-based 3D reconstructions[8]. KNM-ER 1813: CT scan. The left P3 to M^2 were mirror-imaged and matched into the right tooth row. Landmarks were taken from the right side and mirror-imaged using the midsagittal plane of the palate. OH 13: laser scan of the mandible, courtesy of L. Schroeder. OH 65: CT-based 3D surface of a cast, courtesy of R. Blumenschine and R. Leakey.

Prominent early *Homo* specimens not included: KNM-ER 992, 1470, 3733 and UR501 (insufficiently preserved), KNM-ER 730 (*in-vivo* infill and resorption of incisor alveoli), KNM-ER 1805 (maxilla with severe post-mortem distortion; mandible with deformed symphyseal area, likely associated with *in-vivo* loss of anterior dentition), KNM-WT 15000 (developmentally too young), OH 24 (insufficiently preserved and plastic deformation), and D2700 (multiple casts inconclusive regarding poorly preserved alveolar margins and palatal preservation).

Analyses were done using the full hominid sample as well as just the hominins (Fig. 2, Extended Data Figs 4 and 5 and Supplementary Note 2). This approach assesses the fossil hominins with and without the context of the strong morphological contrast between modern humans and great apes.

Procrustes superimposition and principal component analysis. As several fossil specimens were reconstructed using mirror-imaging, we symmetrized all specimens using reflected relabelling[30,32]. The symmetrized landmark coordinates were then converted to shape variables using generalized least squares Procrustes superimposition[33]. This superimposition standardizes position, orientation and scale; the resulting Procrustes coordinates were analysed using multivariate statistics. We used principal component analysis to reduce the dimensions of the high-dimensional data set. Landmarks on the maxilla and mandible were analysed separately. All geometric morphometric and statistical analyses were performed in Mathematica (Wolfram Research) using scripts written by P.G.

Statistical prediction of OH 7 maxillary arcade. So as to compare the dental arcade shape of OH 7 with relevant fossil specimens for which only maxillae exist, we estimated the shape of its maxillary arcade shape using two approaches: an anatomical prediction on the basis of crown occlusion, described below, as well as a statistical model[30] dependent on the covariation of the upper and lower jaw (Extended Data Fig. 8). For the latter we computed five multiple multivariate regression models based on extant great apes and humans to predict the maxillary landmarks from mandibular shape: one regression model pooling all extant species, as well as separate regression models for the extant species. We used principal component analysis to reduce the dimensions of the predictor variables; the regressions were computed in the subspace of the first seven principal components, which together explain more than 95% of the total sample variation. We used multiple reconstructions of the OH 7 maxillary dental arcade to assess the reconstruction uncertainty: one occlusal prediction and five statistical predictions (extant species pooled, *H. sapiens*, *P. troglodytes*, *G. gorilla*, and *Pongo* sp., respectively) for both mandibular reconstructions of OH 7. Our analyses therefore use 11 estimates of the OH 7 maxilla.

We assessed the reliability of our statistical approach by computing a maxillary prediction for every extant mandible on the basis of a cross-validated multiple multivariate regression model (pooling all extant species). A comparison of the actual maxillae to the predicted shapes (Extended Data Fig. 9) shows that the shape of the maxillary arcade can be predicted with high accuracy from the mandibular landmarks. When we applied the extant regression models to the two reconstructions of the OH 7 mandible the predicted maxillae look very similar to one another and

to the anatomical prediction dependent on crown occlusion (Extended Data Fig. 8m). These predicted maxillary shapes differ mostly in the size of the anterior dentition, reflecting the differences among the reference species. The overall arcade shapes of all 11 predictions, however, are almost identical.

Comparisons within and between groups. In order to provide a framework for interpreting the shape differences between fossil specimens we computed Procrustes distances between all possible individual pairs within extant species as a net dissimilarity measure for quantifying the large-scale shape differences. The pairwise Procrustes distances between early *Homo* fossil specimens are plotted in the context of frequency distributions of pairwise Procrustes distances among specimens intraspecifically and interspecifically (Fig. 2b, c, e). The 5% and 95% boundaries of the intraspecific distributions are plotted as vertical lines in the respective group colour.

In Fig. 2b, e and Extended Data Figs 4 and 5 fossils are not assigned group membership a priori. On the basis of these plots it is possible to assess the likelihood that the shape difference between two fossils can be found between any two humans, chimpanzees, orang-utans, or gorillas, and how it compares to the interspecific differences between any two representatives of these extant taxa. Pairwise comparisons of maxillary shape differences include the three predictions of the OH 7 maxilla (Fig. 2e) and, given separately, all five predictions based on individual hominid taxa (Extended Data Fig. 4f). These analyses do not make any assumptions whether a particular fossil is 'typical' for its species or an outlier. They are powerful in demonstrating major species differences, but similarity in arcade shape does not imply conspecificity unless other diagnostic morphology is consistent with the results.

In Fig. 2c we compared the frequency distributions of pairwise Procrustes distances within extant hominid mandibles to within-group distances for early *Homo* samples pooling different combinations of specimens attributed to *H. habilis* (OH 7, OH 13, KNM-ER 1802), *H. erectus* (Dmanisi specimens D2600, D2735, D211) and *H. rudolfensis* (KNM-ER 1482, KNM-ER 60000). For fossils the Procrustes distances between each pair of specimens are given, as well as smooth kernel histograms. For comparison we also simulated a mixed species sample by pooling *Pan troglodytes* and *Gorilla gorilla*; this illustrates the wide within-group distribution of a large mixed-species sample. As a yardstick for assessing the shape variation in fossil groupings we plotted the upper limit of the *G. gorilla* distribution as a dotted black line. Pooling different combinations of specimens attributed to *H. habilis*, *H. erectus* and *H. rudolfensis* yields within-group shape distances that far exceed those found in gorillas, the most variable extant apes.

We stress that our conclusions rest on the Procrustes distance within groups (that is, species). Here variation due to size allometry and neutral evolutionary divergence between populations is expected to be more important than variation due to adaptive divergence between populations (specimens belonging to the same biological species share the same functional adaptation).

Size and allometry. To assess the static allometric effects of jaw size on the shape of the dental arcade, we computed principal component (PC) analyses for both maxillary and mandibular dental arcades in Procrustes form space (Extended Data Fig. 7), and multivariate regressions of upper and lower jaw shape on the respective natural logarithms of centroid size within extant species. The statistical significance of the regression was tested using a permutation test based on the explained variance. In particular, we wanted to explore whether within extant species large mandibles are associated with a narrow dental arcade shape and more parallel post-canine tooth rows than in small mandibles (OH 7 and KNM-ER 1802 have large mandibles). In the Procrustes form space plots (Extended Data Fig. 7a–d) large jaws have low PC scores, and small jaws have high PC scores. As expected, the highly sexually dimorphic gorillas and orang-utans are more variable along PC 1 in Procrustes form space. For maxillary dental arcade shape allometry explains 4% of the total sample variation in *H. sapiens*, 9% in *P. troglodytes*, 18% in *G. gorilla* and 25% in *Pongo* sp. (all $P < 0.001$). For mandibular dental arcade shape allometry accounts for 2% of the total sample variation in *H. sapiens*, 5% in *P. troglodytes*, 15% in *G. gorilla* and 21% in *Pongo* sp. (all $P < 0.001$). In all species, these allometric effects are driven by the size of the canine (Extended Data Fig. 7e–l). These species differences in the fractions of variance explained by static allometry reflect the amount of sexual dimorphism. As expected, this percentage is smallest in humans, and largest in gorillas and orang-utans. Importantly, however, the static allometric effects of jaw size on arcade shape are negligible in all extant species.

Occlusal prediction of the OH 7 maxillary dental arcade. The maxillary dentition of KNM-ER 1590 was chosen as the best match in both size and morphology for the mandibular dentition of OH 7. The total mesiodistal crown length of the left C to M^2 is 612 mm in KNM-ER 1590 (ref. 34) and 592 mm in OH 7 (ref. 1). Silicone moulds were taken of the original teeth of KNM-ER 1590 (left I^1, right and left C, P^3, P^4, M^1, left M^2), and cast in dental stone by one of us (MCD). A micro-CT based surface model of the left M^2 was mirror-imaged to obtain the right M^2, and duplicate M^2s were used as surrogate M^3s. The right I^1 was adapted and modelled in wax from the preserved left I^1, and I^2s were modelled, in proportion and size, on those of OH 16 and KNM-WT 15000 to match the I^1s of KNM-ER 1590. The reconstruction of the OH 7 mandible was stereolithographically printed at 103% to obtain a good size match with the KNM-ER 1590 teeth.

A cast of the OH 7 mandibular dental arcade was mounted in an adjustable dental articulator (*Hanau).* The maxillary dentition was positioned one by one in edge-to-edge (anterior teeth) or centric occlusion (posterior teeth) using the vertical incisal pin to define the midline. Master casts were made of the completed maxillary dental arcade, first in epoxy-resin and then in dental stone. The latter was micro-CT scanned, and a surface model was reduced to 97% to obtain the correct size for the OH 7 mandible (Extended Data Fig. 6c).

Reconstruction of the OH 7 parietals. On the basis of the CT data a currently misaligned anterior fragment on the right parietal was restored to the published configuration[1] (Extended Data Fig. 10). Subsequently, we aligned the mirror-imaged right parietal fragment to the larger left side using the following criteria: (a) superimposition of left and mirror-imaged right asterion and mastoid angle; (b) alignment of the temporal lines on the left and mirror-imaged right parietal fragments; and (c) formation of a continuous sagittal suture. We mirror-imaged the thereby reconstructed left side to the right and aligned both sides. Vault curvature along the sagittal suture and remnants of the right parietal along the suture constrain this alignment and thereby the biparietal width of the resulting reconstruction (Extended Data Fig. 10c). For a second reconstruction (Extended Data Fig. 10d) we realigned the individual pieces that make up the left and right parietal sides. Cracks between the pieces were thereby diminished and the endocranial surfaces of the left and the mirror-imaged right parts could be better aligned, mostly because of the resulting changes in coronal curvature. Compared with the first reconstruction this version is narrower anteriorly and slightly broader posteriorly. Both reconstructions seem valid when different natural anatomical asymmetries are assumed. It has been suggested that the right parietal part, which preserves the occipital angle, could in fact be a left parietal of a second individual, preserving the frontal angle[12]. However, this suggestion is incompatible with the presence of a temporal line that turns inferiorly just anterior to the lambdoid suture (Fig. 3a). Measurements of the parietal reconstructions are given in Supplementary Note 5.

Estimation of the ECV of OH 7. To predict the ECV of OH 7 we quantified parietal size and shape of the two anatomical reconstructions as well as a comparative sample of extant hominids, and used regression-based and thin-plate-spline (TPS) geometric reconstruction methods.

Quantification & sample. On the basis of a published endocranial landmark set[35], we defined 68 3D (semi-)landmarks located on the parietal endocranial surface of OH 7 (Extended Data Fig. 10e). The same subset was retrieved from a sample of 257 extant hominids (84 *H. sapiens*[35], 55 *P. troglodytes*[36], 54 *G. gorilla*[37] and 64 *Pongo* sp[37]). Among these are 124 juveniles with at least M1 fully erupted to bracket the late juvenile age of OH 7. To validate our approach, we used several fossils with known ECVs (*Australopithecus africanus*: Sts 5, Sts 60, Sts 71, StW 505, MLD 37/38; *Homo habilis*: KNM-ER 1813; *Homo erectus*: KNM-ER 3733, KNM-ER 3883, KNM-WT 15000, OH 9). To gain point-to-point correspondence, semi-landmarks were allowed to slide to the landmark configuration on OH 7 (ref. 38). Subsequently we transformed the resulting coordinates into Procrustes shape variables[39,40] and computed Procrustes form variables that capture both shape and size information[41].

Regression model. On the basis of the extant sample, we established a linear regression model (adjusted $R^2 = 0.9908$) to estimate ECV from parietal endocranial form variables. The predicted ECV for the extant hominids is highly correlated to the actual ECV without a taxon-dependent bias towards over- or underestimation (Extend Data Fig. 10f). This also applies to the fossil specimens that were not part of the sample to establish the regression model.

TPS reconstructions. Assuming that each individual of the extant sample had only the parietal area preserved as in OH 7, we predicted the shape of its complete endocast based on TPS warping of all other individuals, and measured the ECV of the resulting 256 reconstructions for each individual. The predicted ECV (the mean of multiple estimates) is highly correlated to the actual ECV (Extended Data Fig. 10g). While there is less discrepancy of predicted versus actual ECV for humans as compared to the regression-based estimation method, the ECV of some species, especially gorillas, is consistently underestimated.

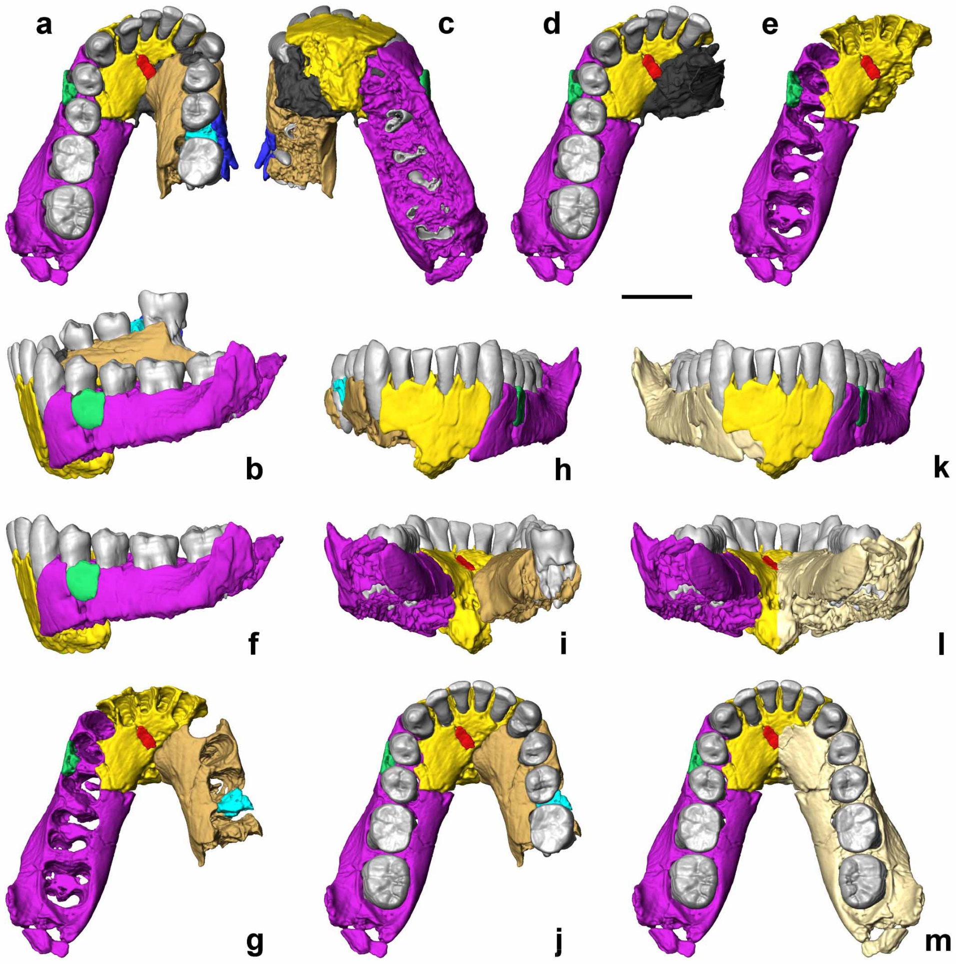
ECV estimates of OH 7. For both anatomical reconstructions of OH 7, we predicted the ECV based on the regression model and the multiple endocast reconstructions generated by TPS-warping. Given that only the parietals are preserved, the ECV estimates inevitably have high uncertainty, reported here as 95% single-prediction bands (regression) and the range of multiple estimates[42] (TPS reconstructions). For the first anatomical reconstruction, the regression-based and TPS-reconstruction-based estimates are 790 ml (678–921 ml) and 771 ml (663–907 ml), respectively, and for the second anatomical reconstruction the estimates are 729 ml (630–844 ml) and 824 ml (710–967 ml), respectively. We consider brain growth completed in individuals with the second molar and canine in occlusion[43,44], and do not make corrections to obtain an adult value (contrary to ref. 1). We consider these values more

reliable than previous ones[1,14–20]: the (semi-)landmark-based quantification of parietal form is more accurate than the individual chords, arcs or volumes used previously; all data were taken from digital copies of the original specimens (avoiding problems of shrinkage or distortion of casts); the comparative sample is large and taxonomically diverse; particularly the regression-based method is not taxon-dependent; and methods were validated for fossil hominins with known ECVs.

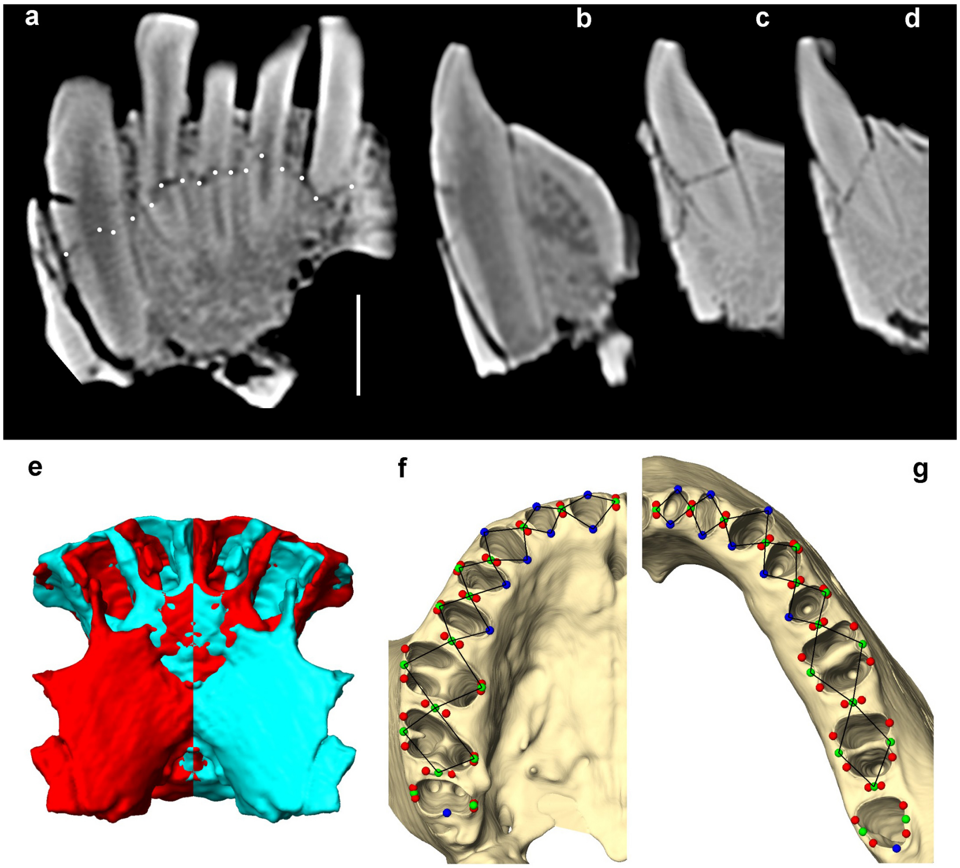
ECV comparison of OH 7, OH 13 & KNM-ER 1813. *Sample. Homo sapiens* 1509 (refs 35, 45–47, Lewis J. E. & Meyer M., personal communication, 2014; worldwide populations). *Pan troglodytes* 184 (refs 48–50, McCarthy R., personal communication, 2005). *Gorilla gorilla* 101 (ref. 48, new data, McCarthy R., personal communication, 2005). *Pongo pygmaeus* 111 (refs 37, 48). *Pongo abelli* 34 (refs 37, 48). The human brain-weight data from ref. 47 were converted to millilitres, by correcting for the density of fresh brain tissue (1.036 g ml^{-1} following ref. 48).

Pairwise comparisons within groups. In light of the large ECV differences between OH 7 (796 ml; 684–927 ml) and the potentially conspecific specimen KNM-ER 1813 (509 ml, ref. 19) we tested whether such ECV differences could be sampled from recent *H. sapiens*, *P. troglodytes*, *P. pygmaeus*, *P. abelli* and *G. gorilla*. We compared ECVs of all possible pairs of individuals within extant groups, and divided the smaller by the larger value; the corresponding frequency distributions and their 5% boundaries are shown in Fig. 3c. This figure also compares OH 7 and OH 13 (650 ml, ref. 20), as well as OH 13 and KNM-ER 1813.

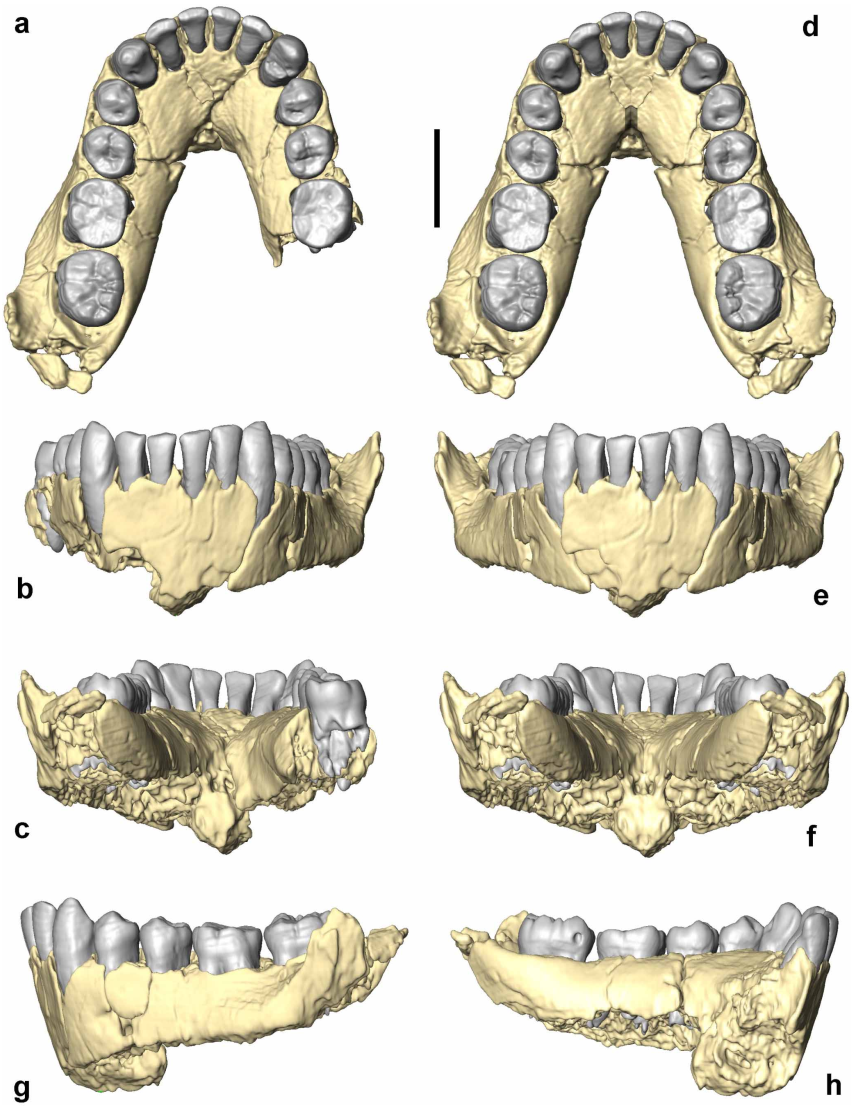
29. Clarke, R. J. A *Homo habilis* maxilla and other newly-discovered hominid fossils from Olduvai Gorge, Tanzania. *J. Hum. Evol.* (2012).
30. Gunz, P., Mitteroecker, P., Neubauer, S., Weber, G. W. & Bookstein, F. L. Principles for the virtual reconstruction of hominin crania. *J. Hum. Evol.* **57,** 48–62 (2009).
31. Kimbel, W. H. & Rak, Y. The cranial base of Australopithecus afarensis: new insights from the female skull. *Phil. Trans. R. Soc. B* **365,** 3365–3376 (2010).
32. Mardia, K. V., Bookstein, F. L. & Moreton, I. J. Statistical assessment of bilateral symmetry of shapes. *Biometrika* **87,** 285–300 (2000).
33. Rohlf, F. J. & Slice, D. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst. Zool.* **39,** 40–59 (1990).
34. Wood, B. A. *Koobi Fora Research Project, vol. 4. Hominid Cranial Remains*. (Clarendon Press, 1991).
35. Neubauer, S., Gunz, P. & Hublin, J. J. The pattern of endocranial ontogenetic shape changes in humans. *J. Anat.* **215,** 240–255 (2009).
36. Neubauer, S., Gunz, P. & Hublin, J. J. Endocranial shape changes during growth in chimpanzees and humans: a morphometric analysis of unique and shared aspects. *J. Hum. Evol.* **59,** 555–566 (2010).
37. Scott, N., Neubauer, S., Hublin, S. & Gunz, P. A shared pattern of postnatal endocranial development in extant hominoids. *Evol. Biol.* **41,** 572–594 (2014).
38. Gunz, P., Mitteroecker, P. & Bookstein, F. L. in *Modern Morphometrics in Physical Anthropology* (ed. Slice, D. E.) 73–98 (Kluwer Academic/Plenum Publishers, 2005).
39. Gower, J. C. Generalized Procrustes analysis. *Pyschometrika* **40,** 33–51 (1975).
40. Rohlf, F. J. & Slice, D. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst. Zool.* **39,** 40–59 (1990).
41. Mitteroecker, P., Gunz, P., Bernhard, M., Schaefer, K. & Bookstein, F. L. Comparison of cranial ontogenetic trajectories among great apes and humans. *J. Hum. Evol.* **46,** 679–697 (2004).
42. Neubauer, S., Gunz, P., Weber, G. W. & Hublin, J. J. Endocranial volume of *Australopithecus africanus*: new CT-based estimates and the effects of missing data and small sample size. *J. Hum. Evol.* **62,** 498–510 (2012).
43. Leigh, S. R. Brain growth, life history, and cognition in primate and human evolution. *Am. J. Primatol.* **62,** 139–164 (2004).
44. Leigh, S. R. Brain size growth and life history in human evolution. *Evol. Biol.* **39,** 587–599 (2012).
45. Lewis, J. E. *et al.* The mismeasure of science: Stephen Jay Gould versus Samuel George Morton on skulls and bias. *PLoS Biol.* **9,** e1001071 (2011).
46. Brown, P. Australian & Asian Palaeoanthropology; Research resources. http://www.peterbrown-palaeoanthropology.net (2014).
47. Marchand, F. *Über das Hirngewicht des Menschen*. (Teubner, Leipzig, 1902).
48. Isler, K. *et al.* Endocranial volumes of primate species: scaling analyses using a comprehensive and reliable data set. *J. Hum. Evol.* **55,** 967–978 (2008).
49. Zuckerman, S. Age-changes in the chimpanzee, with special reference to growth of brain, eruption of teeth, and estimation of age; with a note on the Taung ape. *Proc. Zool. Soc. Lond.* **1,** 1–42 (1928).
50. Neubauer, S., Gunz, P., Schwarz, U., Hublin, J.-J. & Boesch, C. Brief communication: endocranial volumes in an ontogenetic sample of chimpanzees from the Taï Forest National Park, Ivory Coast. *Am. J. Phys. Anthropol.* **147,** 319–325 (2012).
51. van der Merwe, N. J., Masao, F. T. & Bamford, M. K. Isotopic evidence for contrasting diets of early hominins *Homo habilis* and *Australopithecus boisei* of Tanzania. *S. Afr. J. Sci.* **104,** 153–155 (2008).



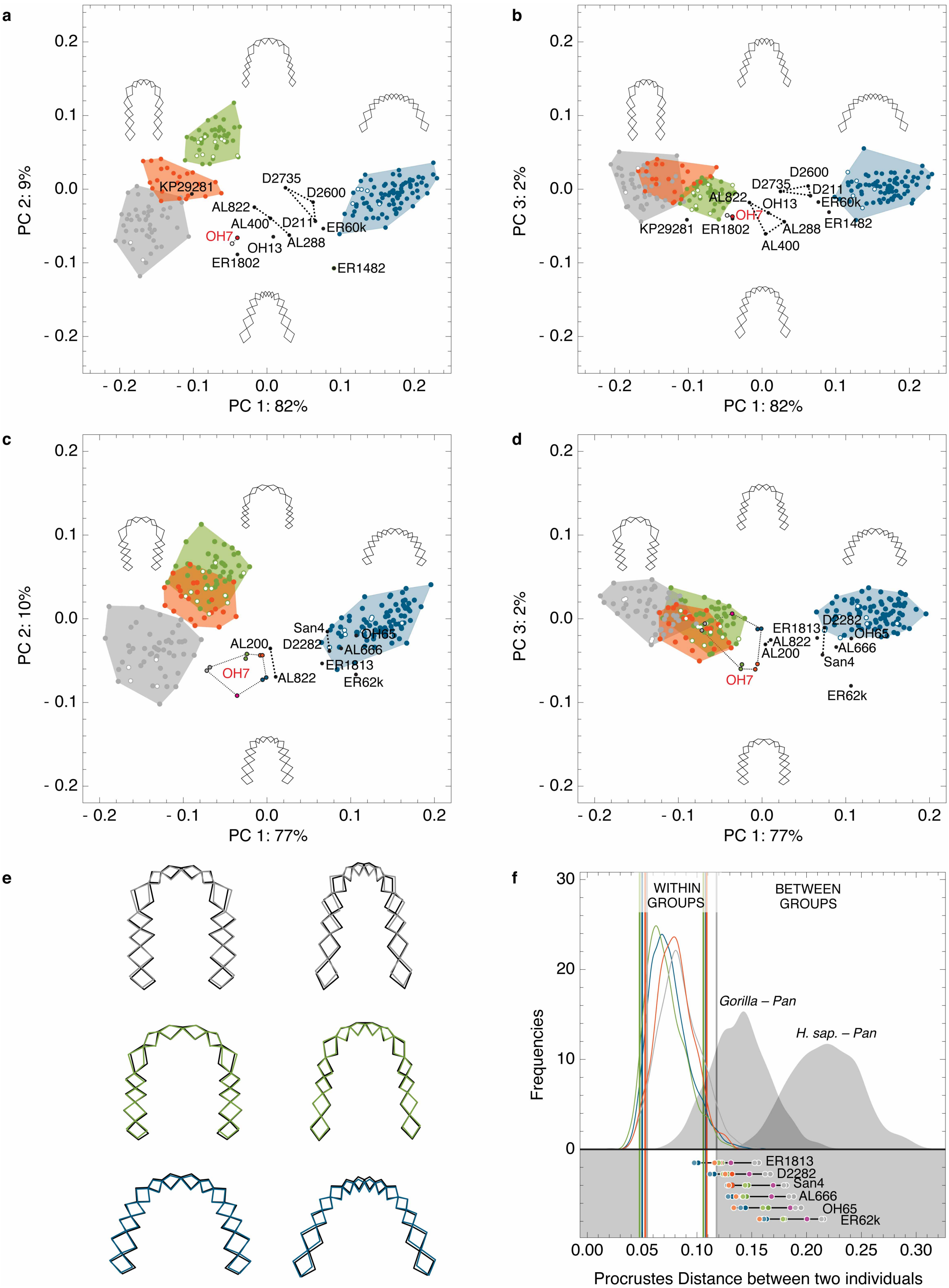
Extended Data Figure 1 | Reconstruction of the OH 7 mandible. **a**–**e**, As preserved in occlusal view (**a**), left lateral view (**b**), inferior view (**c**), without the right corpus, showing the matrix fill (**d**), without the right corpus, matrix and dentition (**e**). **f**, Reconstruction in left lateral view without the right corpus. **g**, Occlusal view of corpus only. **h**–**j**, Full reconstruction in anterior view (**h**), posterior view (**i**) and occlusal view (**j**). **k**–**m**, Reconstruction using the mirror-imaged left corpus in anterior view (**k**), posterior view (**l**) and occlusal view (**m**). Parts are colour-coded as described in the Supplementary Methods. Scale bar is 2 cm.



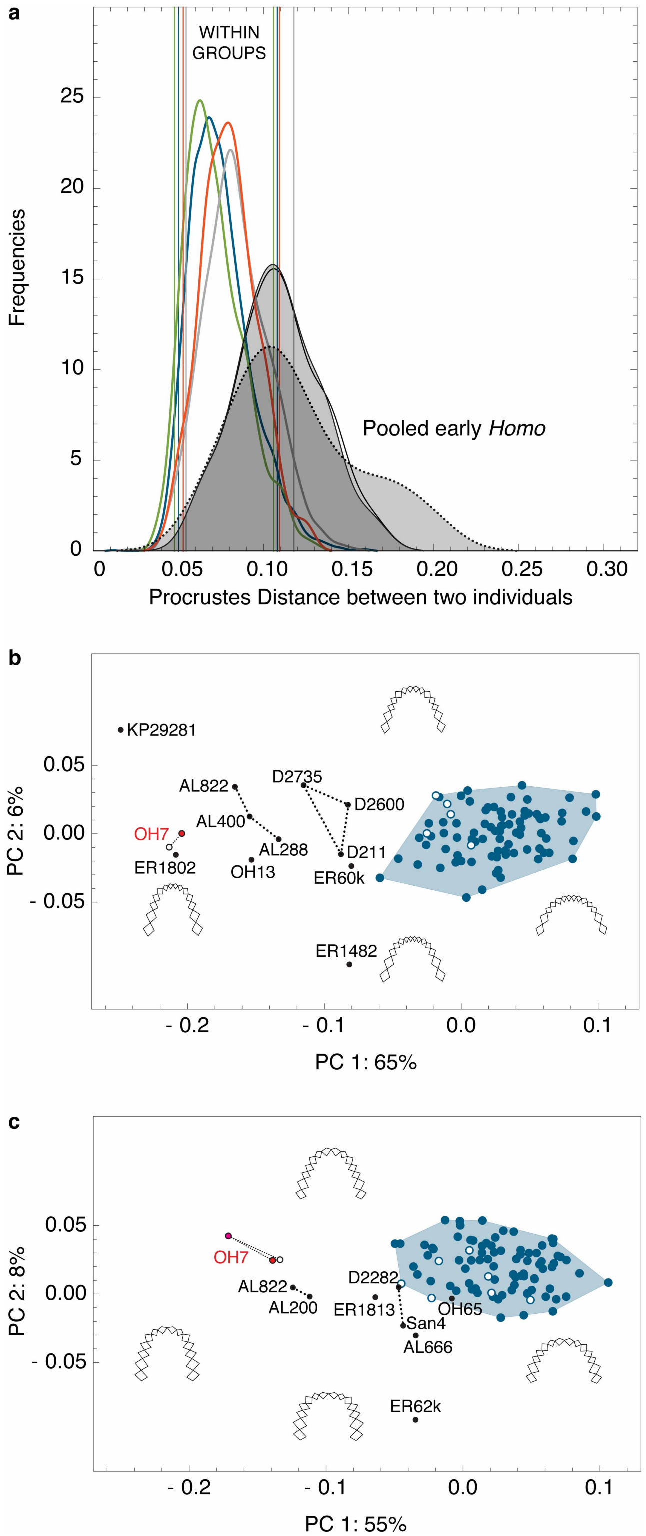
Extended Data Figure 2 | The OH 7 anterior corpus and landmarks of the dental arcade. **a**, Coronal CT section of the anterior corpus showing, from left to right, the roots of the left C to the right I_2, as well as the course of the irregular transverse fracture marked by white dots. **b**–**d**, Labiolingual CT sections through the left C (**b**), I_2 (**c**) and I_1 (**d**). All three show the fracture, and **b** shows the open C root and closed apical alveolar wall. **e**, Anterior corpus as preserved (red) compared with a copy mirror-imaged across the midsagittal plane used for the reconstruction (cyan). The canine alveoli and the anterior symphyseal surface of both sides match well, indicating a lack of overall plastic deformation. The plastically deformed interalveolar septa of the incisors reflect the dislocation of these teeth. Scale bar for **a**–**e** is 1 cm. **f**, **g**, Right maxilla (**f**) and mandible (**g**) of a modern human, showing the landmark positions. Full data sets include both sides of the arcade. Blue and red landmarks were taken from the specimens. The green landmarks were obtained by averaging pairs of red ones, and these were analysed with the blue ones to represent the arcade from left to right M2 (wireframe).



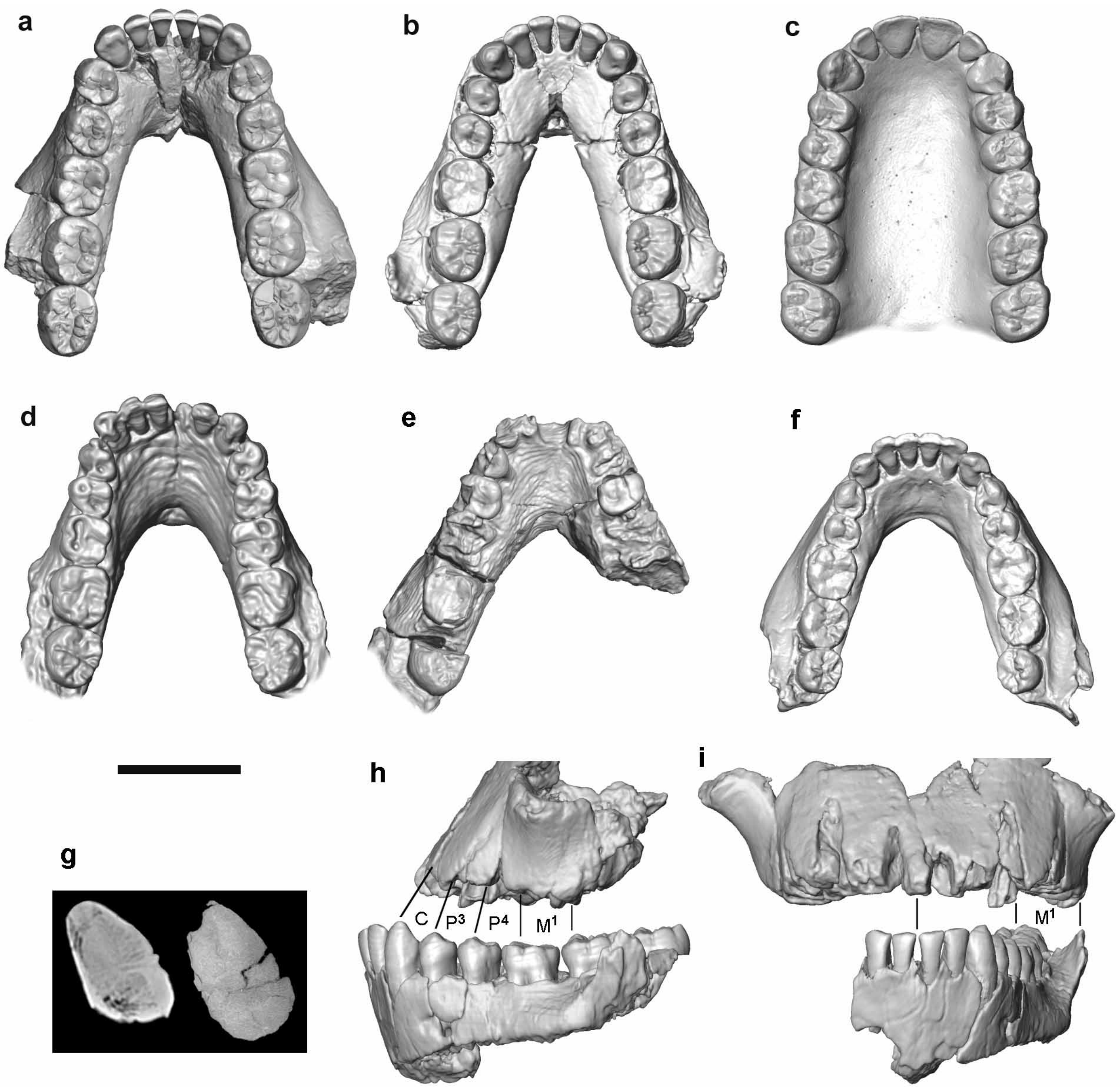
Extended Data Figure 3 | The reconstructed OH 7 mandible. **a**–**c**, Reconstruction using the right corpus in occlusal (**a**), anterior (**b**) and posterior (**c**) view. **d**–**f**, Reconstruction on the basis of mirror-imaged left corpus in occlusal (**d**), anterior (**e**) and posterior (**f**) view. **g**, **h**, Reconstruction without right corpus in left lateral view (**g**) and left medial view (**h**). The hole visible on the lingual crown face of the left M_2 was made to sample tooth enamel for isotope studies[51]. Scale bar is 2 cm.



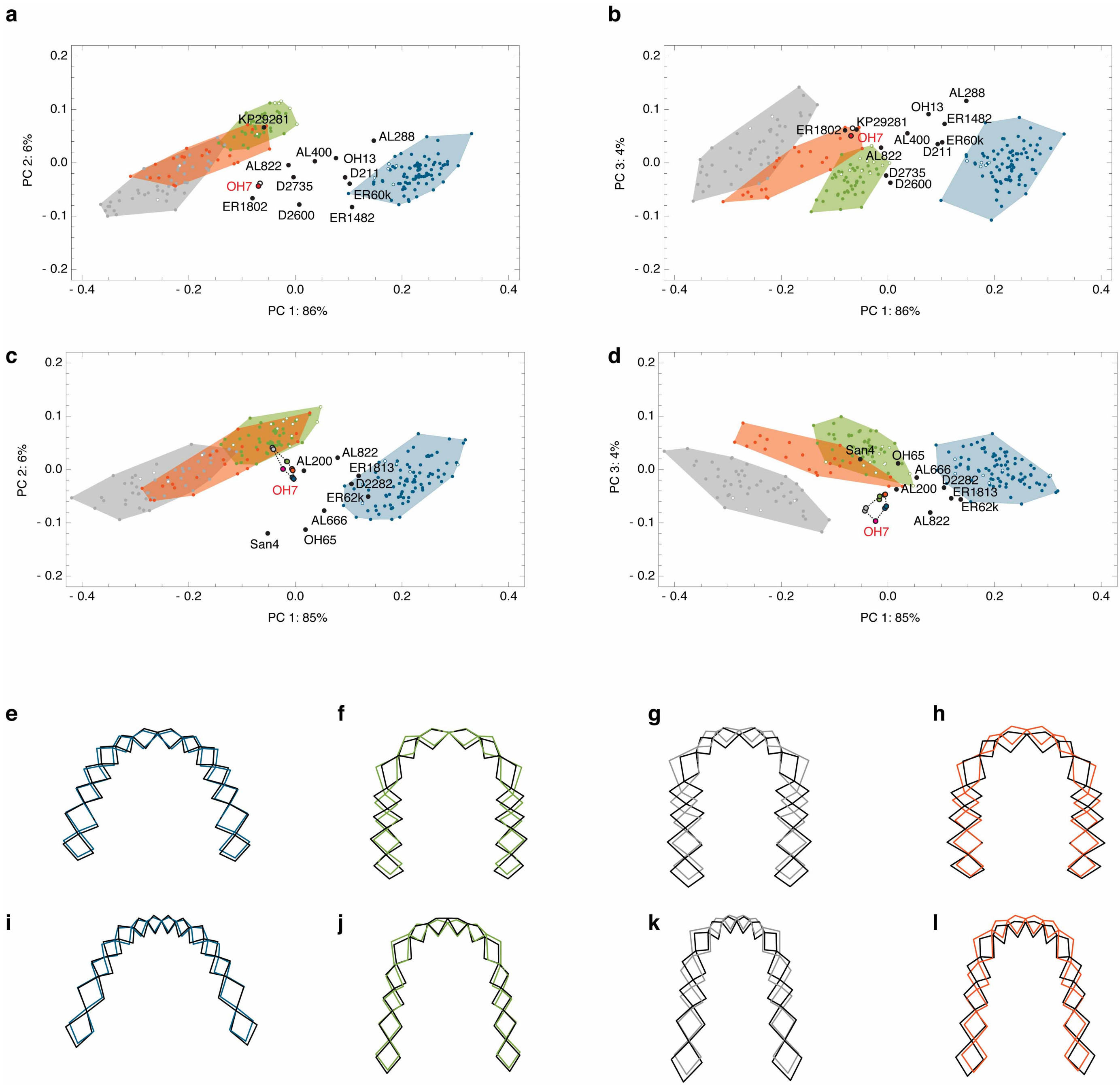
Extended Data Figure 4 | Shape analysis of mandibular and maxillary dental arcades. Group colour codes as in Fig. 2. **a**–**d**, Principal component analysis showing plots of mandibular dental arcade (**a**, PC 1 and PC 2; **b**, PC 1 and PC 3), and maxillary dental arcade (**c**, PC 1 and PC 2; **d**, PC 1 and PC 3). The convex hulls of the extant samples are given, with late juveniles shown as open circles. Wireframes show the shape changes associated with the respective PC axes three standard deviations away from the mean. **e**, Superimposed mean shapes of late juveniles (black) and adults (group colour); maxillary (left) and mandibular (right) dental arcade of *Gorilla gorilla*, *Pan troglodytes* and *Homo sapiens*. **f**, Pairwise Procrustes distances between the nine different maxillary reconstructions of OH 7 and other early *Homo* maxillae: eight statistical predictions are plotted in the colour of the reference species, one occlusal prediction (purple). Frequency plots of the Procrustes distances between all possible individual pairs within (group colour) and between groups (grey); lines represent the 5% limits and 95% limits of these distributions, respectively.



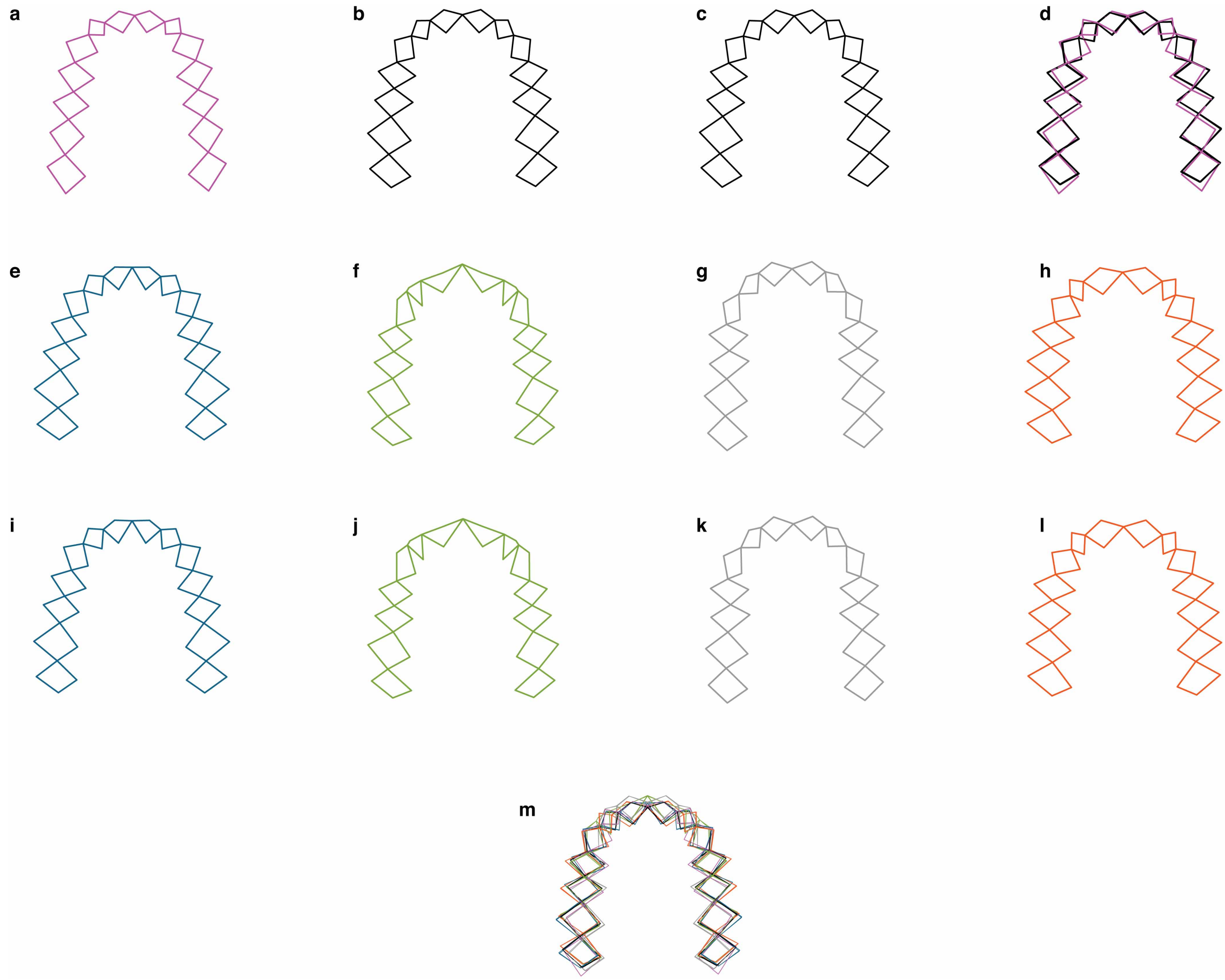
Extended Data Figure 5 | Shape analysis of mandibular and maxillary dental arcades. **a**, Frequency plot of the maxillary shape differences between all pairs within extant groups (colours as in Fig. 2) and within the pooled sample of early *Homo* fossils (black solid lines using statistical predictions of the OH 7 maxilla, dotted line using the occlusal prediction). **b–c**, Principal component analyses showing plots of the mandibular (**b**) and maxillary (**c**) dental arcades. Recent *Homo sapiens* are plotted in blue; late juveniles are shown as open circles. Wireframes show the shape changes associated with the respective principal component axes three standard deviations away from the mean. The red and open circles represent the two alternative mandibular reconstructions of OH 7 (**b**), and the respective statistical predictions of the maxillary dental arcade (**c**). The OH 7 reconstruction of the maxillary dental arcade based on occlusion is plotted in purple (**c**).



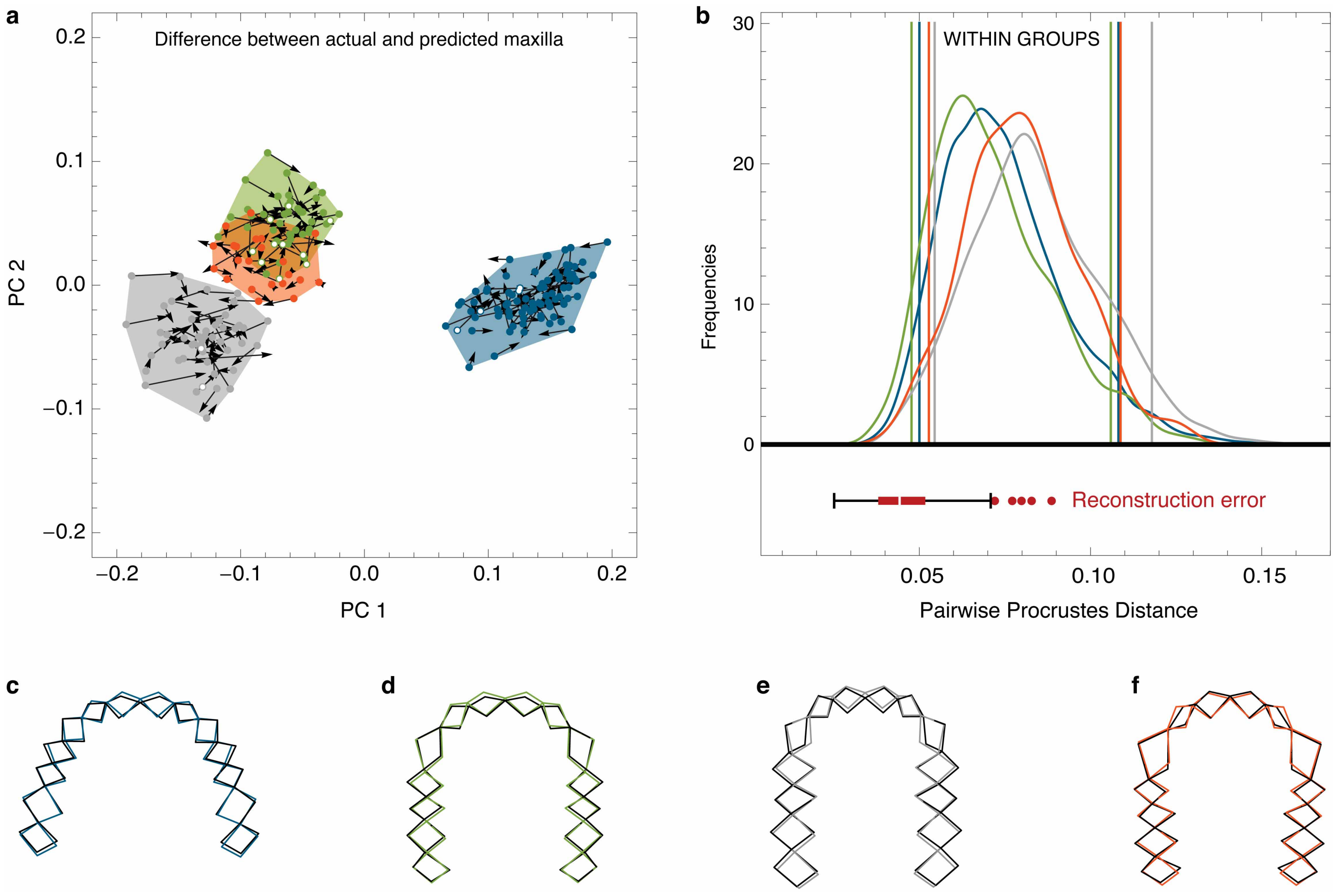
Extended Data Figure 6 | Comparisons of OH 7 mandible. **a**–**f**, Occlusal view of KNM-ER 1802 (**a**), OH 7 (**b**), OH 7 maxillary dental arcade (occlusal prediction) (**c**), A.L 400-1 (**d**), KNM-ER 1482 (**e**) and D211 (**f**). **g**, Midsagittal CT sections of the symphyses of OH 7 (left) and KNM-ER 1802 (right) showing similar ovoid cross-sectional shapes. **h**, Left lateral view of KNM-ER 1470 and OH 7, aligning the reliably identifiable M1 crown position with the corresponding part of the OH 7 row. For C to P4 lines link alveolar margins and corresponding crown position along the OH 7 row. **i**, Anterior view aligning KNM-ER 1470 and OH 7 by their midsagittal plane. KNM-ER 1470 is marked by a shorter C-M^1 row, and a non-projecting anterior row but a wider dental arcade, shown by the position of its left M^1 alveolus well lateral to the corresponding part of OH 7. Scale bar is 3 cm.



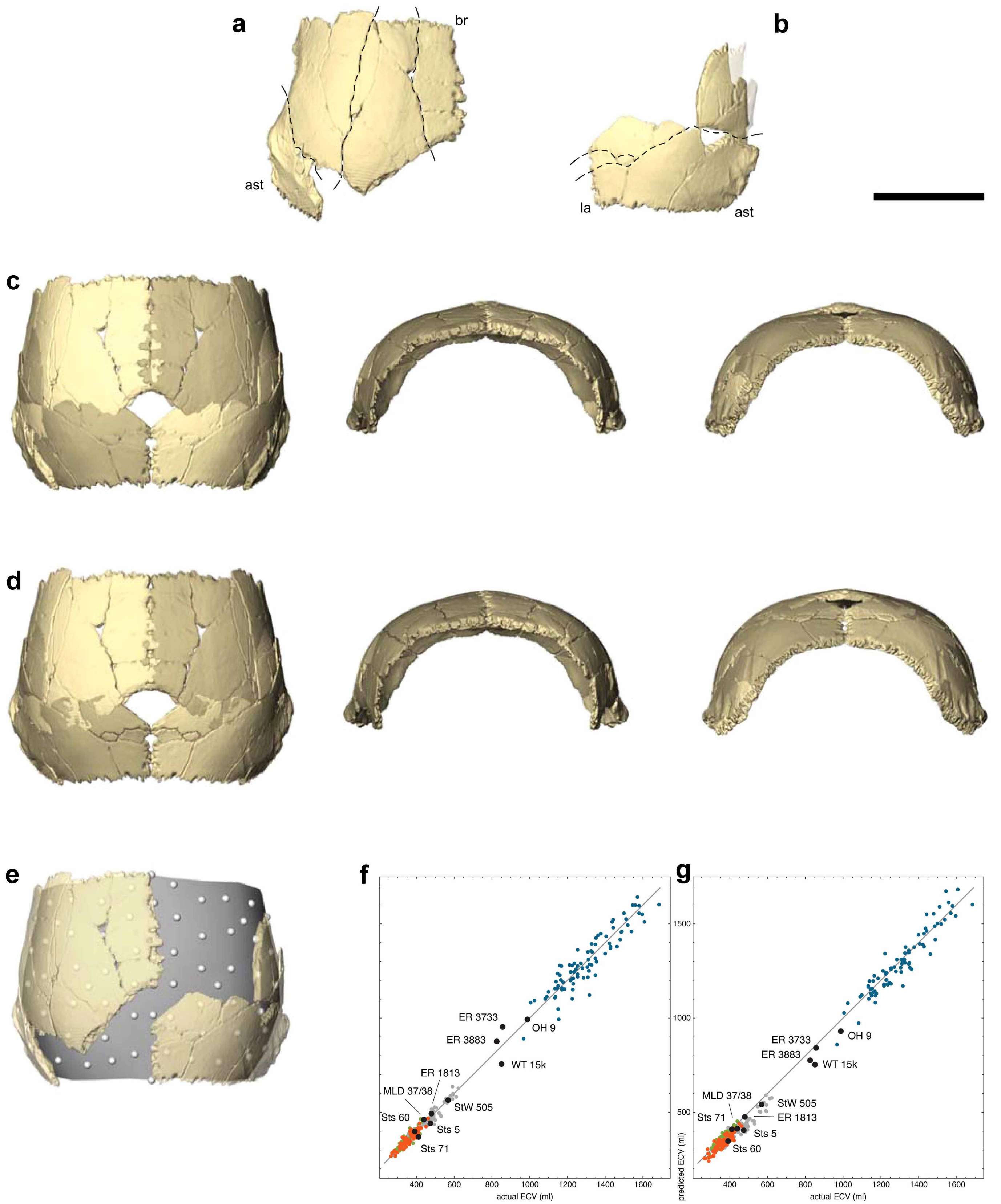
Extended Data Figure 7 | Allometry and dental arcade shape. **a**–**d**, Principal component analysis in Procrustes form space of mandibular dental arcade (**a**, PC 1 and PC 2; **b**, PC 1 and PC 3) and maxillary dental arcade (**c**, PC 1 and PC 2; **d**, PC 1 and PC 3). Colour codes as in Fig. 2. A multivariate regression model was used to assess the covariation of dental arcade shape with size (log centroid size) within extant groups. **e**–**l**, Wireframes show shape predictions for smallest (black) and largest (group colour) centroid size for upper (**e**–**h**) and lower (**i**–**l**) jaw of *H. sapiens* (**e**, **i**), *P. troglodytes* (**f**, **j**), *G. gorilla* (**g**, **k**) and *Pongo* (**h**, **l**). The allometric effects of jaw size on arcade shape are negligible.



Extended Data Figure 8 | Wireframes of maxillary dental arcades. **a**, Predicted OH 7 maxillary arcade, based on dental occlusion. **b**, **c**, Statistical predictions for OH 7 maxilla based on two alternative mandibular reconstructions using a regression model based on all extant species. **d**, Three maxillary predictions of OH 7 (**a**–**c**) superimposed. **e**–**l**, Statistical predictions of OH 7 maxillary dental arcade based on separate regression models for each extant species (applied to two alternative reconstructions of the OH 7 mandible). **e**, **i**, *H. sapiens*; **f**, **j**, *P. troglodytes*; **g**, **k**, *G. gorilla*; **h**, **l**, *Pongo* sp. **m**, All predictions of the OH 7 maxillary dental arcade superimposed.



Extended Data Figure 9 | Reconstruction uncertainty of the maxillary dental arcade. **a**, For every extant specimen we predicted the shape of the maxillary arcade from the mandibular landmarks using a multivariate regression model. Arrows show the difference between the actual and the predicted maxilla in the space of the first two principal components. Colour codes as in Fig. 2. **b**, Box and whisker chart of the Procrustes distances between original and predicted maxillae in the context of the pairwise Procrustes distances within extant groups. Shape differences between original and prediction are usually much smaller than typical within-group shape differences. Even the five outliers with the largest 'reconstruction errors' fall well within the range of shape differences within groups. **c**–**f**, Representative examples of the actual (black) versus predicted (group colour) maxilla wireframes, given for one individual each of: **c**, *Homo sapiens*; **d**, *Pan troglodytes*; **e**, *Gorilla gorilla*; **f**, *Pongo* sp. These predictions use a regression model based on all extant species together.



Extended Data Figure 10 | OH 7 parietal reconstructions and ECV estimation. **a**, **b**, Left and right parietals, with bregma (br), lambda (la) and asterion (ast) indicated. Dashed lines demarcate pieces that were realigned for the second reconstruction. The transparent part of right parietal represents the current, incorrect alignment, corrected for both reconstructions. **c**, First anatomical reconstruction in (left to right) superior, anterior and posterior view, combining the left and right side without realignment of smaller parts (mirror-imaged pieces in darker shade). **d**, Second anatomical reconstruction in (left to right) superior, anterior and posterior view, based on additional realignment of smaller parts. **e**, Endocranial (semi-)landmarks quantifying parietal form. Scale bar, 5 cm. **f**, Regression-based ECV estimates plotted against actual values; grey line indicates perfect match between predicted and actual ECVs; group colours as in Fig. 3 and fossils in black. **g**, Predicted ECVs from TPS reconstructions plotted against actual values.

doi:10.1038/nature14264

AAV-expressed eCD4-Ig provides durable protection from multiple SHIV challenges

Matthew R. Gardner[1]*, Lisa M. Kattenhorn[2]*, Hema R. Kondur[1], Markus von Schaewen[3], Tatyana Dorfman[1], Jessica J. Chiang[2], Kevin G. Haworth[4], Julie M. Decker[5], Michael D. Alpert[2,6], Charles C. Bailey[1], Ernest S. Neale Jr[2], Christoph H. Fellinger[1], Vinita R. Joshi[1], Sebastian P. Fuchs[7], Jose M. Martinez-Navio[7], Brian D. Quinlan[1], Annie Y. Yao[2], Hugo Mouquet[8,9], Jason Gorman[10], Baoshan Zhang[10], Pascal Poignard[11], Michel C. Nussenzweig[8,12], Dennis R. Burton[11,13], Peter D. Kwong[10], Michael Piatak Jr[14], Jeffrey D. Lifson[14], Guangping Gao[15], Ronald C. Desrosiers[2,7], David T. Evans[16], Beatrice H. Hahn[5], Alexander Ploss[3], Paula M. Cannon[4], Michael S. Seaman[17] & Michael Farzan[1]

Long-term *in vivo* expression of a broad and potent entry inhibitor could circumvent the need for a conventional vaccine for HIV-1. Adeno-associated virus (AAV) vectors can stably express HIV-1 broadly neutralizing antibodies (bNAbs)[1,2]. However, even the best bNAbs neutralize 10–50% of HIV-1 isolates inefficiently (80% inhibitory concentration (IC_{80}) $> 5\ \mu g\ ml^{-1}$), suggesting that high concentrations of these antibodies would be necessary to achieve general protection[3–6]. Here we show that eCD4-Ig, a fusion of CD4-Ig with a small CCR5-mimetic sulfopeptide, binds avidly and cooperatively to the HIV-1 envelope glycoprotein (Env) and is more potent than the best bNAbs (geometric mean half-maximum inhibitory concentration (IC_{50}) $< 0.05\ \mu g\ ml^{-1}$). Because eCD4-Ig binds only conserved regions of Env, it is also much broader than any bNAb. For example, eCD4-Ig efficiently neutralized 100% of a diverse panel of neutralization-resistant HIV-1, HIV-2 and simian immunodeficiency virus isolates, including a comprehensive set of isolates resistant to the CD4-binding site bNAbs VRC01, NIH45-46 and 3BNC117. Rhesus macaques inoculated with an AAV vector stably expressed 17–77 $\mu g\ ml^{-1}$ of fully functional rhesus eCD4-Ig for more than 40 weeks, and these macaques were protected from several infectious challenges with SHIV-AD8. Rhesus eCD4-Ig was also markedly less immunogenic than rhesus forms of four well-characterized bNAbs. Our data suggest that AAV-delivered eCD4-Ig can function like an effective HIV-1 vaccine.

Rhesus macaques inoculated with an AAV-based gene-therapy vector express antibody-like immunoadhesins for years, and these immunoadhesins afforded partial protection from a neutralization-sensitive simian immunodeficiency virus (SIV)[2], suggesting that long-term sterilizing protection from HIV-1 might be achievable without a conventional vaccine. Full-length AAV-expressed bNAbs also protected humanized mice from an HIV-1 challenge[1,7]. However, a large fraction of HIV-1 isolates remain partially or wholly resistant to even the best bNAbs, with IC_{80} values greater than $5\ \mu g\ ml^{-1}$ measured under optimal *in vitro* conditions[3–6] (Extended Data Table 1). Higher concentrations will probably be necessary for broad-based protection *in vivo*, but primate studies suggest that these concentrations will be difficult to establish in humans[2,8]. An effective AAV-based vaccine may therefore require broader and more potent inhibitors of HIV-1 entry.

The breadth of an antibody depends on the conservation of its epitope. The two most conserved epitopes of HIV-1 Env are its CD4- and coreceptor-binding sites[9–11]. The immunoadhesin form of CD4, CD4-Ig, has been extensively studied as a therapeutic. It neutralizes most isolates, irreversibly inactivates Env, and is demonstrated safe for use in humans[12–15]. However, its affinity for Env is lower than those of bNAbs[16], and its potency is further compromised by its parallel ability to promote infection[17]. Mimetics of the primary HIV-1 coreceptor CCR5, in particular peptides based on its tyrosine-sulfated amino terminus, have also been characterized[18,19]. These sulfopeptides bind Env specifically but with low affinity in the absence of CD4, in part because they include hydrophobic residues and *O*-linked glycosylation that impede their association with Env[18,20]. CCR5mim1, a 15-amino-acid sulfopeptide derived from the HIV-1 neutralizing antibody E51 (ref. 21), lacks these interfering elements (Fig. 1a) and binds Env with higher affinity than CCR5-based peptides[20,22]. Reflecting the conservation of the sulfotyrosine-binding pockets of Env[9,10], CCR5mim1 binds both CCR5- and CXCR4-dependent Env proteins from all HIV-1 clades[20,22].

We reasoned that a fusion of CD4-Ig and CCR5mim1 would bind Env cooperatively and with higher avidity than either molecule alone. Accordingly, three fusion proteins were generated (sequences in Extended Data Fig. 1). CCR5mim1 was inserted at either the CD4-Ig amino terminus (fusion 1), between the CD4 and Fc domain (fusion 2), or at the CD4-Ig carboxy terminus (fusion 3, renamed eCD4-Ig). All three CD4-Ig variants neutralized CCR5- and CXCR4-dependent isolates more efficiently than did CD4-Ig, with eCD4-Ig consistently the most potent (Extended Data Fig. 2a, b). eCD4-Ig neutralized a wider panel of HIV-1 isolates and SIVmac316 with 10- to 100-fold lower IC_{50} values than CD4-Ig (Fig. 1b). Improved neutralization of SIVmac316 is consistent with conservation of the sulfotyrosine-binding pockets of Env[9,10], and a first indication of the exceptional breadth of eCD4-Ig.

To understand better the markedly greater potency of eCD4-Ig relative to CD4-Ig, we compared their abilities to bind cell-surface-expressed Env trimers (Fig. 1c). At low concentrations, eCD4-Ig bound these trimers more efficiently than did CD4-Ig. Surprisingly, eCD4-Ig saturated trimer-expressing cells with approximately one-third less bound protein than CD4-Ig, suggesting that the sulfopeptides of eCD4-Ig made some CD4-binding sites inaccessible. eCD4-Ig also less efficiently promoted HIV-1 infection of CCR5-positive, CD4-negative cells than CD4-Ig (Fig. 1d), presumably because its sulfopeptides blocked virion access to cell-surface CCR5. Heterodimers of CD4-Ig and eCD4-Ig[23]

[1]Department of Infectious Diseases, The Scripps Research Institute, Jupiter, Florida 33458, USA. [2]Department of Comparative Pathology, Harvard Medical School, New England Primate Research Center, Southborough, Massachusetts 01772, USA. [3]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. [4]Department of Molecular Microbiology and Immunology, Keck School of Medicine of the University of Southern California, Los Angeles, California 90033, USA. [5]Departments of Medicine and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [6]Immunathon Inc., Cambridge, Massachusetts 02141, USA. [7]Department of Pathology, University of Miami Miller School of Medicine, Miami, Florida 33136, USA. [8]Laboratory of Molecular Immunology, The Rockefeller University, New York, New York 10065, USA. [9]Department of Immunology, Institut Pasteur, Paris, 75015, France. [10]Vaccine Research Center, National Institutes of Health, Bethesda, Maryland 20892, USA. [11]Department of Immunology and Microbial Science, IAVI Neutralizing Antibody Center, and Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, The Scripps Research Institute, La Jolla, California 92037, USA. [12]Howard Hughes Medical Institute, New York, New York 10065, USA. [13]Ragon Institute of MGH, MIT and Harvard, Cambridge, Massachusetts 02139, USA. [14]AIDS and Cancer Virus Program, Leidos Biomedical Research, Incorporated, Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702, USA. [15]Gene Therapy Center, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA. [16]Department of Pathology and Laboratory Medicine, University of Wisconsin, Madison, Wisconsin 53711, USA. [17]Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA.
*These authors contributed equally to this work.

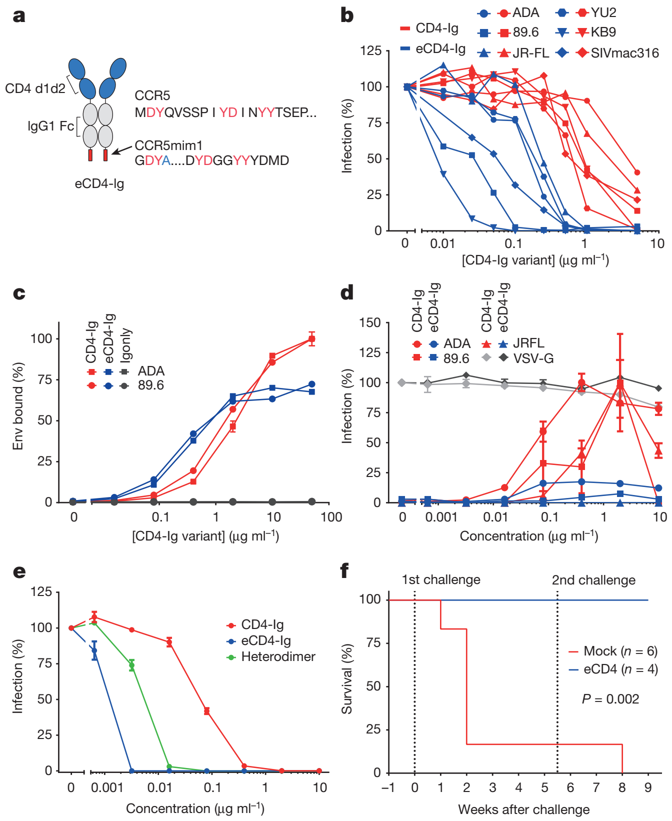


Figure 1 | Functional characterization of eCD4-Ig. **a**, CD4-Ig is comprised of CD4 domains 1 and 2 (blue) fused to the human IgG1 Fc domain (grey). In eCD4-Ig, the sulfopeptide CCR5mim1 (red) is fused to the C terminus of CD4-Ig. The sequence of the CCR5 N terminus is provided for comparison. Common residues, including four CCR5 sulfotyrosines, are shown in red. CCR5mim1 Ala 4 (blue) is substituted with Tyr in CCR5mim2, described below. **b**, HIV-1 pseudotyped with the Env proteins of the indicated HIV-1 or SIV isolates was incubated with GHOST-CCR5 cells and varying concentrations of CD4-Ig (red) or eCD4-Ig (blue). Infection was measured as green fluorescent protein (GFP)-expression by flow cytometry. Errors of replicates are less than 20% of indicated values but not indicated for clarity. **c**, 293T cells transfected to express 89.6 or ADA Env proteins were incubated with the indicated concentrations of CD4-Ig (red), eCD4-Ig (blue) or IgG (grey) and analysed by flow cytometry. **d**, HIV-1 expressing luciferase and pseudotyped with the Env proteins of the indicated isolates was incubated with Cf2Th-CCR5 cells in the presence of varying concentrations of CD4-Ig (red) or eCD4-Ig (blue). Experiment was controlled with HIV-1 pseudotyped with the VSV-G protein (grey). Infection normalized to the maximum value observed for each pseudovirus. **e**, HIV-1 pseudotyped with the 89.6 Env was incubated with TZM-bl cells and varying concentration of CD4-Ig (red), eCD4-Ig (blue) or a CD4-Ig/eCD4-Ig heterodimer (green). Similar experiments using additional Env proteins are shown in Extended Data Fig. 2c, d. **f**, Infection curves of humanized NSG mice with 2–4 μg ml^{-1} of serum eCD4-Ig at time of HIV-1$_{NL4-3}$ challenges (blue line, $n = 5$), or mock treated (red line, $n = 6$) are shown. Three uninfected eCD4-Ig treated mice and the sole uninfected mock treated mouse were rechallenged 5 weeks after the first challenge. Significant protection ($P = 0.002$; Mantel–Cox test) was observed in the eCD4-Ig-treated group. Viral load measurements are shown in Extended Data Fig. 2h. Experiments in **b**–**e** were performed at least twice with each indicated isolate with similar results. Errors bars denote one s.e.m. of duplicates.

neutralized less potently than eCD4-Ig (Fig. 1e and Extended Data Fig. 2c–e), indicating that both eCD4-Ig sulfopeptides engage the Env trimer, consistent with a model of eCD4-Ig bound to Env (Extended Data Fig. 3) and previous studies of CCR5mim1 (ref. 24). Thus, the markedly greater potency of eCD4-Ig relative to CD4-Ig is due in part to the higher avidity with which it binds Env and to its decreased ability to promote infection.

We next assessed eCD4-Ig under more physiological conditions. We observed that eCD4-Ig, but not CD4-Ig, halted replication of infectious viruses in human peripheral blood mononuclear cells (PBMC) at concentrations as low as 125 ng ml^{-1} (Extended Data Fig. 1f, g). We administered sufficient eCD4-Ig to humanized NOD/SCID/*Il2rg*$^{-/-}$ (NSG) mice to maintain serum concentrations of 2–4 μg ml^{-1} at the time of challenge. Five eCD4-Ig-treated mice and six control mice were challenged intravenously with 5×10^4 infectious units of HIV-1$_{NL4-3}$. Five out of six control mice, but no eCD4-Ig-inoculated mice, were infected (Fig. 1f and Extended Data Fig. 2h). Five weeks later, three eCD4-Ig-treated mice and the uninfected control mouse were rechallenged. Again, no eCD4-Ig-treated mouse was infected, whereas the control mouse became infected.

We then characterized the ability of eCD4-Ig to neutralize a diverse panel of neutralization-resistant tier 2 and 3 viruses[25] (Extended Data Figs 4a and 5a). In parallel, we assayed three additional eCD4-Ig variants. In the first, eCD4-Ig^{mim2}, CCR5mim1 was replaced by CCR5mim2, which differs from CCR5mim1 by a single Ala to Tyr substitution[22]. We also introduced a previously characterized Gln 40 to Ala mutation into the CD4 domain 1 of eCD4-Ig (eCD4-Ig^{Q40A})[16]. Both mutations were combined in a final variant (eCD4-Ig^{Q40A,mim2}). eCD4-Ig and these variants substantially outperformed CD4-Ig for every virus in the panel, typically improving neutralization potency by 20- to >200-fold. Underscoring its breadth, eCD4-Ig neutralized SIVmac251 33 times more efficiently than CD4-Ig. In general, the more neutralization-resistant a virus, the better eCD4-Ig and its variants performed relative to CD4-Ig. In most cases, replacement of CCR5mim1 with CCR5mim2 modestly improved neutralization. Similarly, the Gln40Ala mutation also improved neutralization of most HIV-1 isolates, but not of SIVmac251.

We compared eCD4-Ig, eCD4-Ig^{mim2} and eCD4-Ig^{Q40A,mim2} with a panel of 12 antibodies and inhibitors using three additional HIV-1 isolates (Fig. 2a and Extended Data Fig. 6a, b). eCD4-Ig and its variants neutralized the SG3 and YU2 isolates more efficiently than any of these inhibitors. Five bNAbs neutralized JR-CSF more efficiently than any eCD4-Ig variant, but four of these could not neutralize SG3. All eCD4-Ig variants neutralized these isolates with IC$_{50}$ values less than 0.3 μg ml^{-1}, which is more efficiently than CD4-Ig, the tetrameric CD4-Ig variant PRO-542 (refs 12, 14), or the antibodies 2G12, 4E10 and VRC01. eCD4-Ig and its variants, but not three CD4-binding site bNAbs, neutralized the neutralization-resistant SIVmac239 as well as HIV-2 strain ST (Fig. 2b and Extended Data Fig. 6c). As observed with SIVmac251, the Gln40Ala variant was less efficient at neutralizing SIVmac239 and HIV-2. The potency of these eCD4-Ig variants was also reflected in their abilities to mediate antibody-dependent cell-mediated cytotoxicity (ADCC). eCD4-Ig, eCD4-Ig^{mim2} and eCD4-Ig^{Q40A,mim2} each facilitated 30–40 times more killing of infected cells by CD16^{+} natural killer cells[26] than did CD4-Ig or the antibody IgGb12 (Fig. 2c). Thus the C-terminal modification of eCD4-Ig did not interfere with the ADCC effector function of its Fc domain.

We further evaluated eCD4-Ig, eCD4-Ig^{mim2}, eCD4-Ig^{Q40A,mim2} and the bNAb NIH45-46 using nearly every isolate reported to be resistant to either of the CD4bs antibodies NIH45-46 or 3BNC117 (Extended Data Figs 4b and 5b). Both eCD4-Ig variants efficiently neutralized all 38 resistant isolates assayed with IC$_{50}$ values ranging from <0.001 to 1.453 μg ml^{-1}. By contrast, 26 isolates in this panel were confirmed to be resistant to NIH45-46. Previous reports found 29 and 18 isolates to be resistant to 3BCN117 and VRC01, respectively[4,6]. Figure 3 and Extended Data Fig. 7 summarize the neutralization studies compiled from the experiments in Figs 1 and 2 and Extended Data Figs 4–6, and from previous studies of VRC01 and 3BNC117 against the same isolates[4]. They show that the geometric mean IC$_{50}$ and IC$_{80}$ values of eCD4-Ig and its variants are less than 0.05 μg ml^{-1} (500 pM) and 0.2 μg ml^{-1} (2 nM), respectively, roughly 3–4 times lower than those of VRC01, NIH45-46 or 3BNC117. Importantly, our lead eCD4-Ig variant, eCD4-Ig^{mim2}, neutralized 100% of the isolates assayed at concentrations (IC$_{50}$ < 1.5 μg ml^{-1}; IC$_{80}$ < 5.2 μg ml^{-1}) that are probably sustainable in humans.

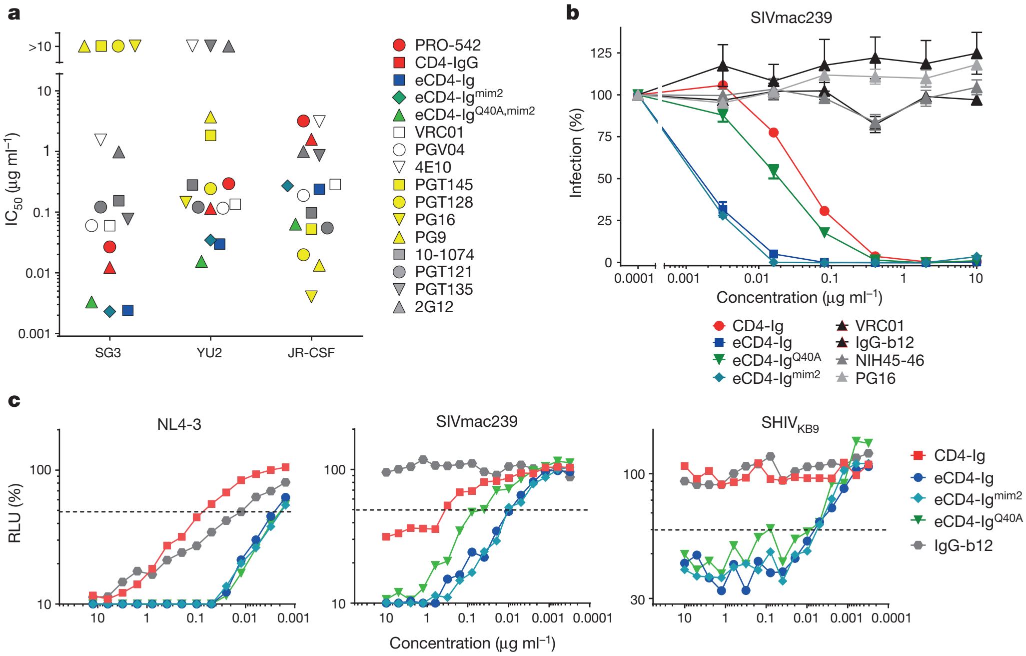


Figure 2 | Comparison of eCD4-Ig variants and HIV-1 neutralizing antibodies. **a**, HIV-1 pseudotyped with the Env proteins of the indicated isolates were incubated with TZM-bl cells and varying concentrations of the indicated entry inhibitors, and the resulting IC_{50} values are plotted. IC_{90} values and standard errors are presented in Extended Data Fig. 6a, b. **b**, Experiments similar to those in **a** except that HIV-1 pseudotyped with the SIVmac239 Env was incubated with varying concentrations of CD4-Ig, eCD4-Ig variants or CD4bs antibodies. Extended Data Fig. 6c shows a similar study using the HIV-2 ST Env. Errors bars denote one s.e.m. of triplicates. **c**, ADCC activity was assessed using CEM.NKR-CCR5 target cells incubated with infectious HIV-1 NL4-3, SIVmac239 or SHIV$_{KB9}$ for 4 days. Cells were then incubated with KHYG-1 NK effector cells[26] for 8 h in the presence of the indicated inhibitors. ADCC activity was measured as loss of luciferase activity from the target cells. RLU, relative light units. All experiments represented in this figure were performed at least twice with each isolate and inhibitor with similar results. Error bars indicate one s.e.m. of triplicates.

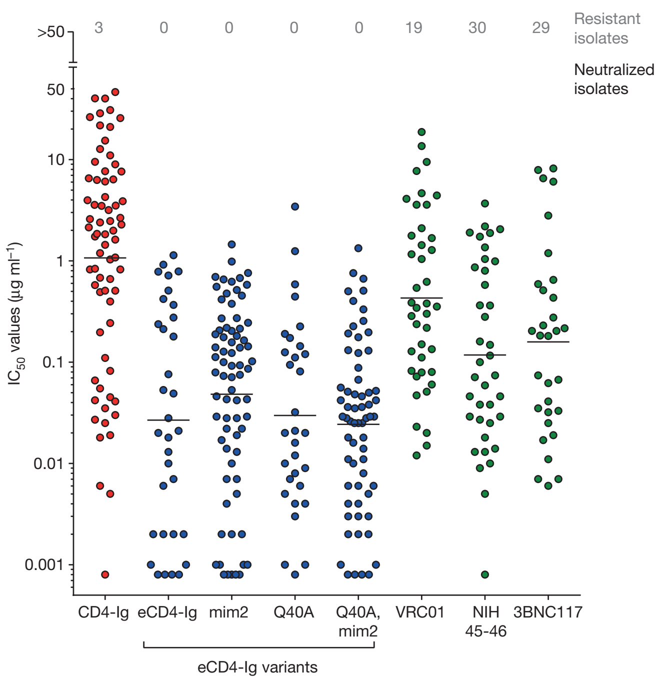


Figure 3 | Summary of HIV-1, HIV-2 and SIV neutralization studies. The IC_{50} values from studies of Figs 1b, 2a, b and Extended Data Figs 4–6 are plotted. The numbers of isolates resistant to 50 μg ml^{-1} of the indicated inhibitors are indicated at the top. Geometric means are calculated for neutralized isolates and indicated with horizontal lines. Note that these data include 38 HIV-1 isolates selected for resistance to NIH45-46 or 3BNC117, so that isolates resistant to these antibodies are over-represented. Nonetheless, the geometric mean values of neutralized viruses are consistent with previous reports (Extended Data Table 1). Data for VRC01 and 3BNC117 were reported previously[4,6]. IC_{80} values are presented in Extended Data Fig. 7.

Finally, using a rhesus macaque form of eCD4-Ig^{mim2}, we investigated whether AAV-delivered eCD4-Ig could function like an HIV-1 vaccine. To minimize potential adverse reactions, the Fc domain of rhesus IgG2, which binds Fc receptors and complement less efficiently than IgG1, was used. We also introduced an Ile39Asn mutation into the CD4 domain[27] to correct partially the lower affinity of rhesus CD4 for most HIV-1 isolates (Extended Data Fig. 8a, b). The gene for the resulting construct, rh-eCD4-IgG2^{I39N,mim2} (described hereafter as rh-eCD4-Ig), was cloned into a single-stranded AAV2 vector (AAV-rh-eCD4-Ig; Extended Data Fig. 8c). A total of 2.5×10^{13} AAV1-encapsidated particles delivering this vector were administered into the quadriceps of four four-year-old male Indian-origin rhesus macaques. To promote rh-eCD4-Ig sulfation, a separate single-stranded AAV vector expressing rhesus tyrosine-protein sulfotransferase 2 (AAV-rh-TPST2; Extended Data Fig. 8c) was co-administered with AAV-rh-eCD4-Ig at a 1:4 ratio. No adverse reactions were observed in any of the AAV-rh-eCD4-Ig-inoculated macaques. These macaques and four age- and gender-matched controls were challenged intravenously with increasing doses of SHIV-AD8 (Fig. 4a, b). Sixteen weeks after AAV inoculation, two control macaques became infected following challenge with 200 pg p27. A subsequent 400 pg challenge infected a third control animal, and, after resisting an additional 400 pg challenge, the final control was infected with 800 pg, 34 weeks from the date of AAV inoculation. None of these challenges infected AAV-rh-eCD4-Ig-inoculated macaques, indicating that eCD4-Ig protected them from four doses capable of infecting control animals.

Measured rh-eCD4-Ig titres in the serum stabilized to between 17 and 77 μg ml^{-1} over the last 10 weeks of the 40-week study period (Fig. 4c). Two macaques expressed less than 20 μg ml^{-1} at the time of the final 800 pg challenge, suggesting that this concentration could prevent many otherwise infectious exposures in humans. Sera from inoculated macaques neutralized HIV-1 as efficiently as laboratory-prepared rh-eCD4-Ig mixed with pre-inoculation sera (Fig. 4d and Extended Data Fig. 8d),

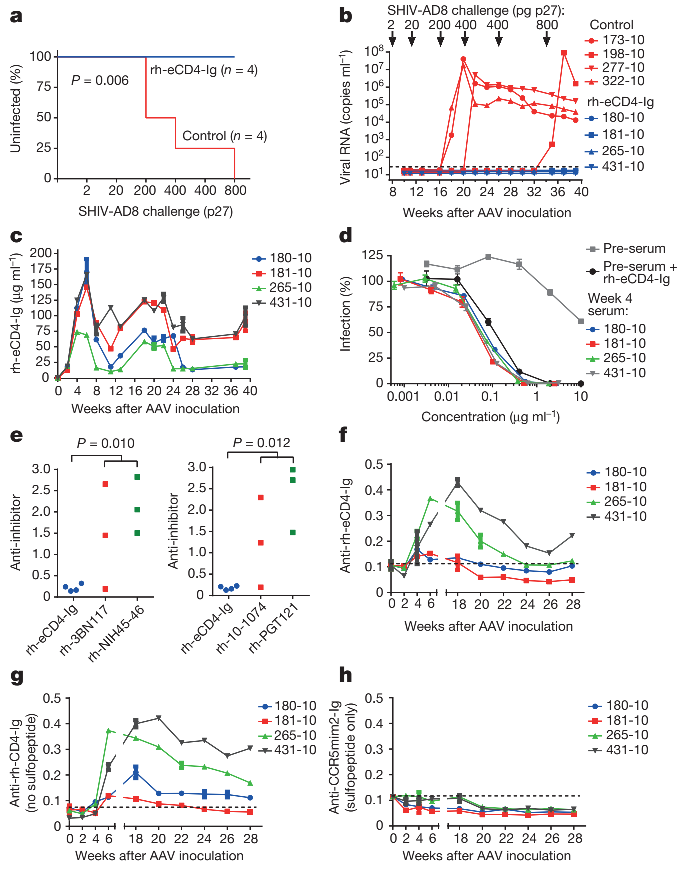


Figure 4 | AAV-rh-eCD4-Ig protects rhesus macaques from SHIV-AD8. **a**, Infection analysis comparing four male Indian-origin rhesus macaques inoculated intramuscularly with 2×10^{13} AAV particles delivering rh-eCD4-Ig (blue) and four age- and gender-matched controls (red). At 8, 11, 16, 20, 26 and 34 weeks after inoculation, macaques were challenged with the indicated p27 titres of SHIV-AD8. Significant protection ($P = 0.006$; Mantel–Cox test) was observed in the AAV-rh-eCD4-Ig-treated group. **b**, Viral loads of inoculated (blue) and control (red) macaques are shown, with the time and titre of challenge indicated above the graph. **c**, Concentrations of rh-eCD4-Ig in the sera of inoculated macaques were measured by ELISA to week 40 after inoculation. **d**, The neutralizing potency of macaque sera obtained 4 weeks after AAV-inoculation was compared to pre-inoculation sera (pre-sera), and pre-sera mixed with laboratory-produced rh-eCD4-Ig, as in Fig. 2b. **e**, Anti-transgene antibody responses in AAV-rh-eCD4-Ig-inoculated macaques were compared to those in macaques inoculated with AAV expressing the indicated bNAbs bearing constant regions of rhesus IgG2. Sera from 4 weeks after inoculation were analysed. Plates were coated with equivalent amounts of rh-eCD4-Ig or rhesus forms of bNAbs and incubated with sera and anti-rhesus lambda chain (left) or kappa chain (right) antibody conjugated to horseradish peroxidase. Note that 3BNC117 and NIH45-46 bear a kappa light chain, whereas PGT121 and 10-1074 bear a lambda light chain, so that only host antibody responses were detected. Values indicate absorbance at 450 nM. P values (Student's two-tailed t-test) are indicated above the figures. **f**, The sensitivity of the assay in **e** was increased to measure longitudinally the anti-rh-eCD4-Ig activity in the sera of inoculated macaques. Both anti-kappa and anti-lambda secondary antibodies were used. Values are scaled for comparison to values in **e**. **g**, **h**, The same assay as in **f** except that responses to rh-CD4-Ig, lacking CCR5mim2 (**g**) or to CCR5mim2 fused to a human IgG1 Fc domain (**h**) were measured. Experiments in **c**–**h** were performed at least twice with similar results. Errors bars denote one s.e.m. of duplicates.

indicating that the eCD4-Ig was efficiently sulfated and fully active *in vivo*. We also compared macaque humoral responses to expressed rh-eCD4-Ig and to four AAV-expressed bNAbs inoculated for a separate study. 3BNC117, NIH45-45, 10-1074 and PGT121, each bearing rhesus IgG2 and light-chain constant domains, elicited markedly higher endogenous antibody responses than did rh-eCD4-Ig, consistent with their high levels of somatic hypermutation (Fig. 4e). To investigate the target of the anti-rh-eCD4-Ig responses, we increased the sensitivity of our assay and compared longitudinally the reactivity of inoculated rhesus sera to a series of antigens. rh-eCD4-Ig (Fig. 4f) and rh-CD4-Ig (without the CCR5mim2 sulfopeptide; Fig. 4g) were recognized by rhesus sera with nearly the same reactivity, whereas CCR5mim2 fused to a human IgG1 Fc domain was not (Fig. 4h), indicating that the sulfopeptide was not immunogenic. Rhesus CD4 domains 1 and 2 fused to a human IgG1 Fc was much less reactive than the same CD4 domains fused to the rhesus IgG2 Fc, without or with the Ile39Asn mutation (Extended Data Fig. 8e, f), whereas an unrelated construct bearing the rhesus IgG2 Fc domain showed no reactivity (Extended Data Fig. 8g), suggesting that a neo-epitope formed by the rhesus CD4 and Fc domains was recognized by most anti-rh-eCD4-Ig antibodies. Thus eCD4-Ig is less immunogenic than bNAbs, and can be expressed for at least 40 weeks at concentrations that are well tolerated and protective against several robust SHIV-AD8 challenges.

A key question is whether eCD4-Ig or a similar construct could be used to prevent new HIV-1 infections in a population, and whether it might do so more effectively than a bNAb. We show that AAV-delivered rhesus eCD4-Ig protected all inoculated macaques from multiple infectious doses that are probably higher than those present in most human transmission events, although we have not yet tested protection from mucosal challenges. Protection lasted at least 34 weeks after inoculation (Fig. 4b), and other studies indicate that these protective titres can be sustained for several years[2]. Previous studies of CD4-Ig indicate that it is safe when passively administered[12,14], and in particular it does not engage MHC II or otherwise interfere with immune function[13], although further safety studies of eCD4-Ig are warranted. eCD4-Ig has fewer non-self B- and T-cell epitopes than heavily hypermutated bNAbs, and thus elicits fewer endogenous antibodies that can impair its expression and activity (Fig. 4e). Its most prominent non-self element is its sulfopeptide, which did not elicit any measurable antibody responses (Fig. 4f–h). However, the clearest advantage of eCD4-Ig over bNAbs is its potency and its unmatched breadth (Fig. 3 and Extended Data Figs 4–7). The breadth of eCD4-Ig arises from the necessary conservation of its binding sites on Env, suggesting that emergence of eCD4-Ig escape variants in a population is less likely than with bNAbs. Moreover, any virus that does bypass prophylaxis is likely to bind CD4 and CCR5 less efficiently in the continued presence of eCD4-Ig, and may therefore be less efficiently retransmitted. Its potency suggests that relatively lower concentrations of eCD4-Ig will be sufficient to protect against most circulating viruses, a feature that may be critical to its use with AAV in humans. Although there are remaining challenges, these observations suggest that AAV-expressed eCD4-Ig could provide effective, long-term and near universal protection from HIV-1.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 June 2013; accepted 27 January 2015.

Published online 18 February 2015.

1. Balazs, A. B. *et al.* Antibody-based protection against HIV infection by vectored immunoprophylaxis. *Nature* **481,** 81–84 (2011).
2. Johnson, P. R. *et al.* Vector-mediated gene transfer engenders long-lived neutralizing activity and protection against SIV infection in monkeys. *Nature Med.* **15,** 901–906 (2009).
3. Diskin, R. *et al.* Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science* **334,** 1289–1293 (2011).
4. Huang, J. *et al.* Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491,** 406–412 (2012).
5. Walker, L. M. *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477,** 466–470 (2011).
6. Scheid, J. F. *et al.* Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333,** 1633–1637 (2011).
7. Lewis, A. D., Chen, R., Montefiori, D. C., Johnson, P. R. & Clark, K. R. Generation of neutralizing activity against human immunodeficiency virus type 1 in serum by antibody gene transfer. *J. Virol.* **76,** 8769–8775 (2002).
8. Greig, J. A. *et al.* Intramuscular injection of AAV8 in mice and macaques is associated with substantial hepatic targeting and transgene expression. *PLoS ONE* **9,** e112268 (2014).

9. Rizzuto, C. D. *et al.* A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science* **280,** 1949–1953 (1998).
10. Huang, C. C. *et al.* Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science* **317,** 1930–1934 (2007).
11. Lagenaur, L. A., Villarroel, V. A., Bundoc, V., Dey, B. & Berger, E. A. sCD4–17b bifunctional protein: extremely broad and potent neutralization of HIV-1 Env pseudotyped viruses from genetically diverse primary isolates. *Retrovirology* **7,** 11 (2010).
12. Fletcher, C. V. *et al.* Nonlinear pharmacokinetics of high-dose recombinant fusion protein CD4-IgG2 (PRO 542) observed in HIV-1-infected children. *J. Allergy Clin. Immunol.* **119,** 747–750 (2007).
13. Hussey, R. E. *et al.* A soluble CD4 protein selectively inhibits HIV replication and syncytium formation. *Nature* **331,** 78–81 (1988).
14. Jacobson, J. M. *et al.* Single-dose safety, pharmacology, and antiviral activity of the human immunodeficiency virus (HIV) type 1 entry inhibitor PRO 542 in HIV-infected adults. *J. Infect. Dis.* **182,** 326–329 (2000).
15. Haim, H. *et al.* Soluble CD4 and CD4-mimetic compounds inhibit HIV-1 infection by induction of a short-lived activated state. *PLoS Pathog.* **5,** e1000360 (2009).
16. Moebius, U., Clayton, L. K., Abraham, S., Harrison, S. C. & Reinherz, E. L. The human immunodeficiency virus gp120 binding site on CD4: delineation by quantitative equilibrium and kinetic binding studies of mutants in conjunction with a high-resolution CD4 atomic structure. *J. Exp. Med.* **176,** 507–517 (1992).
17. Sullivan, N. *et al.* Determinants of human immunodeficiency virus type 1 envelope glycoprotein activation by soluble CD4 and monoclonal antibodies. *J. Virol.* **72,** 6332–6338 (1998).
18. Farzan, M. *et al.* Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. *Cell* **96,** 667–676 (1999).
19. Farzan, M. *et al.* A tyrosine-sulfated peptide based on the N terminus of CCR5 interacts with a CD4-enhanced epitope of the HIV-1 gp120 envelope glycoprotein and inhibits HIV-1 entry. *J. Biol. Chem.* **275,** 33516–33521 (2000).
20. Dorfman, T., Moore, M. J., Guth, A. C., Choe, H. & Farzan, M. A tyrosine-sulfated peptide derived from the heavy-chain CDR3 region of an HIV-1-neutralizing antibody binds gp120 and inhibits HIV-1 infection. *J. Biol. Chem.* **281,** 28529–28535 (2006).
21. Choe, H. *et al.* Tyrosine sulfation of human antibodies contributes to recognition of the CCR5 binding region of HIV-1 gp120. *Cell* **114,** 161–170 (2003).
22. Chiang, J. J. *et al.* Enhanced recognition and neutralization of HIV-1 by antibody-derived CCR5-mimetic peptide variants. *J. Virol.* **86,** 12417–12421 (2012).
23. Ridgway, J. B., Presta, L. G. & Carter, P. 'Knobs-into-holes' engineering of antibody CH3 domains for heavy chain heterodimerization. *Protein Eng.* **9,** 617–621 (1996).
24. Kwong, J. A. *et al.* A tyrosine-sulfated CCR5-mimetic peptide promotes conformational transitions in the HIV-1 envelope glycoprotein. *J. Virol.* **85,** 7563–7571 (2011).
25. Seaman, M. S. *et al.* Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. *J. Virol.* **84,** 1439–1452 (2010).
26. Alpert, M. D. *et al.* A novel assay for antibody-dependent cell-mediated cytotoxicity against HIV-1- or SIV-infected cells reveals incomplete overlap with antibodies measured by neutralization and binding assays. *J. Virol.* **86,** 12039–12052 (2012).
27. Humes, D., Emery, S., Laws, E. & Overbaugh, J. A species-specific amino acid difference in the macaque CD4 receptor restricts replication by global circulating HIV-1 variants representing viruses from recent infection. *J. Virol.* **86,** 12472–12483 (2012).

Acknowledgements This project was supported by National Institutes of Health (NIH) grants R01 AI091476 and R01 AI080324 (M.F.), P01 AI100263 (G.G., R.C.D., M.F.), RR000168 (M.R.G., L.M.K., D.T.E., R.C.D., M.F.), R01 AI058715 (B.H.H.), by the Intramural Research program of the Vaccine Research Center, NIAID, NIH (J.G., B.Z., P.D.K.), and by federal funds from the National Cancer Institute, NIH under contract no. HHSN261200800001E. The authors would like to thank H. Choe and M. Martin for critical advice.

Author Contributions M.R.G. and L.M.K. contributed equally to this work. M.R.G., L.M.K., H.R.K., M.V.S., T.D., J.J.C., M.D.A., M.P., J.D.L., R.C.D., D.T.E., B.H.H., P.M.C., M.S.S., A.P. and M.F. designed experiments. M.R.G., L.M.K., H.R.K., M.V.S., T.D., J.J.C., K.G.H., J.M.D., M.D.A., C.C.B., C.H.F., V.R.J., B.D.Q. and A.Y.Y. performed experiments. L.M.K. conducted all non-human primate studies. J.G. and P.D.K. assisted with modelling. J.M.M.-N., H.M., B.Z., P.P., M.S.S., M.C.N. and D.R.B. contributed advice and critical reagents. M.F. conceived the study and, with important assistance from M.R.G. and L.M.K., wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.F. (mfarzan@scripps.edu).

METHODS

Plasmids and cells. Plasmid expressing CD4-Ig was previously described[20]. Fusion constructs were created by adding sequences encoding CCR5mim1 and tetra-glycine linker to N terminus (fusion1) or between domain 2 and human Fc (fusion2) of CD4-Ig by inverse PCR. eCD4-Ig (fusion3) and eCD4-Ig^{mim2} were created by adding sequence encoding a tetra-glycine linker and CCR5mim1 or CCR5mim2, respectively, to the C terminus of CD4-Ig by inverse PCR. The Gln 40 to Ala mutation was introduced in eCD4-Ig and eCD4-Ig^{mim2} by Quickchange PCR. The eCD4-Ig/CD4-Ig heterodimer was generated as previously described[23] and analysed by SDS–PAGE under reducing and non-reducing conditions. rh-eCD4-Ig, consisting of rhesus CD4 domains 1 and 2 bearing an Ile39Asn mutation, rhesus IgG2 Fc and CCR5mim2, was synthesized and cloned into a previously described single-stranded AAV plasmid[2]. AAV expression plasmids for HIV-1 antibodies were created by synthesizing the variable heavy and light chains of 3BNC117, NIH45-46, PGT121 and 10-1074 with the rhesus heavy and light constant regions, and cloning these genes into a previously described ssAAV plasmid[2]. The following reagent was obtained through the NIH AIDS Reagent Program (Division of AIDS, NIAID, NIH): CMVR-VRC01-H, CMVR-VRC01-L, from J. Mascola[28,29], pNL4-3.Luc.R-.E- from N. Landeau[30,31], TZM-bl cells from J. C. Kappes, X. Wu and Tranzyme Inc[32–36], SF162 gp160 from L. Stamatatos and C. Cheng-Mayer[37], and GHOST-CCR5 and -CXCR4-cells from V. KewalRamani and D. Littman. Human embryonic kidney HEK293T cells were obtained from ATCC. Cf2Th-CD4$^+$.CCR5$^+$ and CfTh-CCR5$^+$ cells were a gift from H. Choe. No testing for mycoplasma contamination was performed in any cell line after their receipt from these contributors. The variable heavy and light chains of IgG-b12, NIH45-46, 3BNC117, 10-1074 and PGT121 were cloned into the CMVR-VRC01-H and -L plasmids. Plasmids encoding TPST-2 or the envelope glycoproteins pNL4-3Δenv, 89.6, ADA, SG3, SA32, YU2, JRFL, KB9, VSV-G, HIV-2 ST, SIVmac239, SIVmac316 and replicative 89.6 or SG3 viruses were previously described[20,21,38–40].

Purification of antibodies, CD4–Ig and eCD4-Ig variants. Production of CD4-Ig, eCD4-Ig variants and antibodies was performed as previously described[41]. In brief, HEK293T cells in 140 mm plates were transfected with 25 μg per plate at 50% confluency by the calcium phosphate transfection method. Plasmids encoding sulfated proteins were cotransfected with a plasmid encoding human tyrosine protein sulfotranserase 2 (TPST2). At 12 h after transfection, 10% FBS-DMEM media was replaced with serum-free 293 Freestyle media (Invitrogen). Media was collected after 48 h, debris was cleared by centrifugation for 10 min at 1,500*g* and filtered using 0.45-μm filter flasks (Millipore). Complete protease inhibitor cocktail (Roche) was added to the filtered supernatants. A 500-μl bed volume of Protein A sepharose beads (GE Healthcare) was added and agitated at 4 °C overnight. The bead–media mixture was collected by gravity flow column (Biorad) and was washed with 30 ml PBS (Lonza) plus 0.5 M NaCl (0.65 M NaCl final) followed by 10 ml PBS. Protein was eluted with 3 M $MgCl_2$ in PBS. Buffer was exchanged for PBS and protein was concentrated to 1 mg ml^{-1} by Ultrafiltration (Amicon Ultra) at 4,000*g*.

Flow cytometry analysis of CD4-Ig and eCD4-Ig binding to cell-expressed envelope glycoprotein. HEK293T cells were transfected with plasmids expressing envelope glycoprotein lacking cytoplasmic residues 732 to 876 (HXBc2 numbering) together with plasmid encoding the tat protein. Transfection medium was replaced after an overnight incubation and cells were collected 48 h after transfection. Collected cells were washed twice in flow cytometry buffer (PBS with 2% goat serum, 0.01% sodium azide). Cells were incubated with CD4-Ig or eCD4-Ig on ice for 1 h and then washed twice with flow cytometry buffer. A secondary antibody recognizing human Fc (Jackson Immuno Research) was added to the cells for 30 min. Cells were washed twice with flow cytometry buffer, twice with PBS, and resuspended in 1% paraformaldehyde solution. Binding was analysed with an Accuri C6 Flow Cytometer (BD Biosciences) and data analysed with the C6 Software (BD Biosciences).

Viral entry enhancement assay. HIV-1 pseudovirus expressing firefly luciferase was pre-incubated with titrated amounts of CD4-Ig or eCD4-Ig variants in DMEM (10% FBS) for 1 h at 37 °C. CD4-negative Cf2Th-CCR5 cells were collected and diluted in DMEM (10% FBS) to 100,000 cells ml^{-1} and added to the pseudovirus/inhibitor mixture. Cells were then incubated for 48 h at 37 °C. Viral entry was analysed using Britelite Plus (Perkin Elmer) and luciferase activity of cell lysates was read using a Victor X3 plate reader (Perkin Elmer).

HIV-1 neutralization assays. GHOST-CCR5 or -CXCR4 cells were plated into 12-well plates at 50,000 cells per well. HIV-1 pseudovirus was diluted in RPMI and titrated amounts of CD4-Ig, fusion1, fusion2 or eCD4-Ig were added. Virus and inhibitor were incubated at room temperature for 20 min and added to the cells for 2 h at 37 °C. Cells were then washed with serum-free medium and then incubated in 1 ml of DMEM (10% FBS) for 48 h at 37 °C. Cells were collected by trypsinization, fixed in 1% paraformaldehyde in PBS, and viral entry was determined by flow cytometry based on GFP expression.

For studies of infectious virus, unstimulated PBMCs were collected and resuspended in RPMI medium (15% FBS, 20 U ml^{-1} IL-2). Cells were plated in a 12-well plate at 10^6 cells per well. HIV-1 was diluted in RPMI and varying amounts of inhibitor were added. The virus and inhibitor was incubated at room temperature for 20 min and added to the cells for 3 h at 37 °C. Cells were then washed with serum-free medium and resuspended in fresh RPMI medium (15% FBS, 20 U ml^{-1} IL-2). At 3-day intervals after infection, supernatants were collected and fresh RPMI medium (15% FBS, 20 U ml^{-1} IL-2) was added to the cells. Supernatants were analysed for viral infection by ELISA with Alliance HIV-1 p24 antigen ELISA kit (Perkin Elmer).

TZM-bl neutralization assays were performed as previously described[42]. In brief, HIV-1 pseudoviruses were pre-incubated with titrated amounts of CD4-Ig or eCD4-Ig variants in DMEM (10% FBS) for 1 h at 37 °C. TZM-bl cells were collected and diluted in DMEM (10% FBS) to 100,000 cells ml^{-1} and added to the pseudovirus/inhibitor mixture. Cells were then incubated for 48 h at 37 °C. Viral entry was analysed using Britelite Plus (Perkin Elmer) and luciferase activity was read using a Victor X3 plate reader (Perkin Elmer). All neutralization and enhancement studies of Figs 1–4 were performed at least twice in triplicate. All error bars represent s.e.m.

Antibody-dependent cell-mediated cytotoxicity assays. ADCC activity was performed as previously described[43]. In brief, CEM.NKR CCR5 CD4$^+$ T cells were infected 4 days with infectious HIV-1 NL4.3, SHIV-KB9 or SIVmac239. After 4 days, KHYG-1 effector cells were co-incubated with infected cells in the presence of titrated CD4-Ig, eCD4-Ig variants, or the b12 antibody for 8 h. ADCC activity was measured by luciferase activity as above.

Production of HIV-1$_{NL4-3}$ stocks and SHIV-AD8-EO stocks for *in vivo* studies. A molecular clone of HIV-1$_{NL4-3}$ was obtained from the AIDS Research and Reference Reagent Program (ARRRP), Division of AIDS, NIAID, NIH from material deposited by S. Gartner, M. Popovic, R. Gallo and M. Martin. Virus stocks were produced in 293T cells by transient transfection using TurboFect (Thermo Scientific) and 12 μg of proviral plasmid. Supernatants were collected at 40 h, filtered through 0.45-μm filters, and dispensed into single use doses and frozen at −80 °C. Viruses were quantified by p24 ELISA (Zeptometrix) and by GHOST cell titer[44] to determine infectious units per millilitre (IU ml^{-1}). Titering was performed per the GHOST cell line protocol obtained through ARRRP. The molecular clone of SHIV-AD8-EO was a gift from M. Martin[45]. 293T cells were plated in 140 mm flasks and transfected with 80 μg DNA per plate by calcium phosphate technique. At 12 h after transfection, flasks were replaced with fresh DMEM (10% FBS). Medium was collected at 48 h after transfection, frozen at −80 °C, and titred using an SIV p27 ELISA kit (ABL).

Haematopoietic stem cell isolation and NSG mouse transplantation. Human CD34$^+$ haematopoietic stem cells were isolated from fetal livers obtained from Advanced Bioscience Resources, INC (ABR). Tissue was disrupted and incubated with 1 mg ml^{-1} collagenase/dispase (Roche Applied Sciences) for 15 min at 37 °C. Cells were isolated by passing the disrupted tissue through a 70-μm filter. Red blood cells were lysed in BD Pharm Lyse (BD Biosciences), with CD34$^+$ cells being isolated using CD34 MACS microbeads (Miltenyi) according to manufacturer's instructions with an additional purification step using a second column. NOD.Cg-Prkdc scid Il2rγ tm1Wj/Szj (NOD/SCID/*Il2rg*$^{-/-}$, NSG) mice were obtained from Jackson Laboratories. Neonatal mice received 150 cGy radiation, and 2–4 h later 1×10^6 CD34$^+$ haematopoietic stem cells in 1% heparin (Celgene) via intrahepatic injection. Mice were monitored for engraftment levels of human CD45$^+$ cells and development of T cells and B cells at 8, 10 and 12 weeks after engraftment.

Mouse infections, treatment and analysis. Humanized mice with evidence of human CD4$^+$ T-cell development in blood were infected with 5×10^4 IU of HIV-1NL4.3 by intraperitoneal injection. Mice were administered with 65 μg of eCD4-Ig once weekly for the first 2 weeks, starting at 8 days before the HIV-1 challenge, and then twice weekly starting at week 3 by retro-orbital injection while under anaesthetization by 2.5% isoflurane. Mock-treated mice received a retro-orbital injection of PBS 1 and 8 days before HIV-1 challenge, and were anaesthetized in parallel with eCD4-Ig mice throughout. Every week after infection the mice were anaesthetized by inhalation of 2.5% isoflurane and blood was collected retro-orbitally for analysis. At week 6, three eCD4-Ig-treated mice and one mock-treated mouse (who had not become infected) were challenged a second time with 5×10^4 IU HIV-1 NL4-3. Mouse blood was blocked for 20 min at room temperature in FBS (Denville) and stained with appropriate antibodies for 15 min at room temperature. Red blood cells were removed by incubation in BD FACS Fix/Lysing Solution (BD Biosciences), which was removed by dilution with PBS before analysis by flow cytometry. HIV-1 levels in peripheral blood were determined by extracting viral RNA from mouse plasma at each blood draw using a viral RNA isolation kit (Qiagen) followed by Taqman One-Step RT–PCR (Life Technologies) using a primer and probe set targeting the HIV-1 LTR region, as previously described[46,47]. Reactions were performed and analysed using a 7500 Fast Realtime PCR System (Life Technologies). To analyse engrafted T cells by flow cytometry, stained cells were acquired on a FACS Canto II (BD Biosciences) and analysed using FlowJo software v7.6.5 (Tree

Star Inc.). Blood samples were stained using human-specific antibodies at a 1:20 dilution for CD4-V450 (RPA-T4), CD8-APC (RPAT8), CD3-PE (UCHT1) and CD45-PerCP (TUI16) (BD Bioscience). Up to 10,000 events were recorded for viable cell populations and gated based on fluorescence minus one controls as previously described[46]. All mouse studies were performed in accordance with the Scripps Research Institute Institutional Review Board, protocol number 14-018. No statistical methods were used to predetermine sample size.

AAV inoculation of rhesus macaques. Eight 4-year-old AAV1-negative male Indian-origin rhesus macaques were housed at the New England Primate Research Center in accordance with standards set forth by the American Association for Accreditation of Laboratory Animal Care. Their weights at the time of AAV inoculation ranged from 5.2 to 8.2 kg. Macaques were separated into age- and weight-matched control groups, but blinding and randomization were not performed. Four macaques were inoculated with 1 ml RPMI containing 2.5×10^{13} AAV1 particles delivering 80% of a single-strand rh-eCD4-Ig transgene (IgG2 isotype) and 20% of a single-strand rhesus TPST-2 transgene into each quadriceps muscle (two 0.5 ml per injections per quadriceps muscle). Five millilitres of sera was obtained every 1–2 weeks after AAV inoculation beginning at week 4. Animals were challenged at week 8 after inoculation with 2 pg p27 of SHIV-AD8-EO. SHIV-negative animals were repeatedly challenged with escalating doses of SHIV-AD8-EO up to 800 pg p27. Plasma viral loads were quantified as previously described[45].

For AAV studies of bNAbs, six 2-year-old AAV1-negative Indian-origin rhesus macaques (two males and four females) were housed at the New England Primate Research Center in accordance with standards set forth by the American Association for Accreditation of Laboratory Animal Care. Three macaques were inoculated with 1 ml RPMI containing 1×10^{13} AAV1 particles delivering single-strand rh-3BNC117-IgG2 transgene into one quadriceps (two 0.5-ml injections) and 1 ml RPMI containing 1×10^{13} AAV1 particles delivering single-stranded rh-10-1074-IgG2 transgene into the contralateral quadriceps (two 0.5-ml injections). The other three macaques were inoculated with 1 ml RPMI containing 1×10^{13} AAV1 particles delivering single-strand rh-NIH45-46-IgG2 transgene into one quadriceps (two 0.5-ml injections) and 1 ml RPMI containing 1×10^{13} AAV1 particles delivering single-strand rh-PGT121-IgG2 transgene into the contralateral quadriceps (two 0.5-ml injections). Five millilitres of sera was obtained every 2 weeks beginning at week 2 and analysed by ELISA. All primate studies were performed in accordance with the Harvard Medical School Standing Committee on Animals protocol number 04888.

eCD4-Ig, rh-eCD4-Ig and anti-transgene antibody concentrations in NSG mice and rhesus macaque sera. *In vivo* concentrations of eCD4-Ig, rh-eCD4-Ig were measured by ELISA as previously described[2]. In brief, to measure NSG mouse and macaque serum concentrations, ELISA plates (Costar) were coated with 5 μg ml^{-1} SIV gp120 overnight at 4 °C. Plates were washed with PBS-T (PBS plus 0.05% Tween-20) twice and blocked with 5% milk in PBS for 1 h at 37 °C. Sera serially diluted in 5% milk in PBS were added to the plate and incubated for 1 h at 37 °C. Samples were washed five times with PBS-T and a horseradish peroxidase secondary antibody (Jackson Immuno Research) recognizing human IgG1. Plates were incubated for 1 h at 37 °C and then washed ten times with PBS-T. TMB solution (Fisher) was added for 10 min at room temperature and then stopped with TMB Stop Solution (Southern Biotech). Absorbance was read at 450 nm by a Victor X3 plate reader (Perkin Elmer) and compared with a standard curve generated using a rh-eCD4-Ig mixed with pre-inoculation sera. Anti-rh-eCD4-Ig antibodies and anti-bNAb antibodies were measured in the same way except that ELISA plates were coated with 5 μg ml^{-1} of various constructs. Constructs so assayed included rh-eCD4-Ig, rh-CD4-Ig^{I39N}, rh-CD4-Ig domains 1 and 2 (with or without Ile39Asn) bearing a human IgG1 Fc and hinge domain, C-terminal CCR5mim2-Ig (human IgG1 Fc and hinge, no CD4 domains), NIH45-46 bearing the rhesus IgG2 Fc domain and hinge, or HIV-1 bNAbs. Serum samples were diluted 10- or 20-fold and blocked in 5% milk in PBS. Anti-transgene antibodies were measured using secondary antibodies detecting either the kappa or lambda light chain (Southern Biotech) that was opposite of the antibody being assayed when comparing the anti-bNAb response to that to rh-eCD4-Ig. Both anti-kappa and anti-lambda secondary antibodies were used when measuring anti-rh-eCD4-Ig responses alone. TMB solution was added for 10–15 min at room temperature and measured as described above.

28. Wu, X. *et al.* Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329,** 856–861 (2010).
29. Barouch, D. H. *et al.* A human T-cell leukemia virus type 1 regulatory element enhances the immunogenicity of human immunodeficiency virus type 1 DNA vaccines in mice and nonhuman primates. *J. Virol.* **79,** 8828–8834 (2005).
30. He, J. *et al.* Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. *J. Virol.* **69,** 6705–6711 (1995).
31. Connor, R. I., Chen, B. K., Choe, S. & Landau, N. R. Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. *Virology* **206,** 935–944 (1995).
32. Platt, E. J., Bilska, M., Kozak, S. L., Kabat, D. & Montefiori, D. C. Evidence that ecotropic murine leukemia virus contamination in TZM-bl cells does not affect the outcome of neutralizing antibody assays with human immunodeficiency virus type 1. *J. Virol.* **83,** 8289–8292 (2009).
33. Takeuchi, Y., McClure, M. O. & Pizzato, M. Identification of gammaretroviruses constitutively released from cell lines used for human immunodeficiency virus research. *J. Virol.* **82,** 12585–12588 (2008).
34. Wei, X. *et al.* Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob. Agents Chemother.* **46,** 1896–1905 (2002).
35. Derdeyn, C. A. *et al.* Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. *J. Virol.* **74,** 8358–8367 (2000).
36. Platt, E. J., Wehrly, K., Kuhmann, S. E., Chesebro, B. & Kabat, D. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *J. Virol.* **72,** 2855–2864 (1998).
37. Harouse, J. M. *et al.* Mucosal transmission and induction of simian AIDS by CCR5-specific simian/human immunodeficiency virus SHIV(SF162P3). *J. Virol.* **75,** 1990–1995 (2001).
38. Choe, H. *et al.* The orphan seven-transmembrane receptor apj supports the entry of primary T-cell-line-tropic and dualtropic human immunodeficiency virus type 1. *J. Virol.* **72,** 6113–6118 (1998).
39. Choe, H. *et al.* The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* **85,** 1135–1148 (1996).
40. Farzan, M. *et al.* A tyrosine-rich region in the N terminus of CCR5 is important for human immunodeficiency virus type 1 entry and mediates an association between gp120 and CCR5. *J. Virol.* **72,** 1160–1164 (1998).
41. Quinlan, B. D., Gardner, M. R., Joshi, V. R., Chiang, J. J. & Farzan, M. Direct expression and validation of phage-selected peptide variants in mammalian cells. *J. Biol. Chem.* **288,** 18803–18810 (2013).
42. Li, M. *et al.* Human immunodeficiency virus type 1 Env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J. Virol.* **79,** 10108–10125 (2005).
43. Alpert, M. D. *et al.* ADCC develops over time during persistent infection with live-attenuated SIV and is associated with complete protection against $SIV_{mac}251$ challenge. *PLoS Pathog.* **8,** e1002890 (2012).
44. Mörner, A. *et al.* Primary human immunodeficiency virus type 2 (HIV-2) isolates, like HIV-1 isolates, frequently use CCR5 but show promiscuity in coreceptor usage. *J. Virol.* **73,** 2343–2349 (1999).
45. Shingai, M. *et al.* Antibody-mediated immunotherapy of macaques chronically infected with SHIV suppresses viraemia. *Nature* **503,** 277–280 (2013).
46. Holt, N. *et al.* Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nature Biotechnol.* **28,** 839–847 (2010).
47. Rouet, F. *et al.* Transfer and evaluation of an automated, low-cost real-time reverse transcription-PCR test for diagnosis and monitoring of human immunodeficiency virus type 1 infection in a West African resource-limited setting. *J. Clin. Microbiol.* **43,** 2709–2717 (2005).
48. Tran, E. E. *et al.* Structural mechanism of trimeric HIV-1 envelope glycoprotein activation. *PLoS Pathog.* **8,** e1002797 (2012).
49. Sauer-Eriksson, A. E., Kleywegt, G. J., Uhlen, M. & Jones, T. A. Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. *Structure* **3,** 265–278 (1995).
50. Huang, C. C. *et al.* Structural basis of tyrosine sulfation and VH-gene usage in antibodies that recognize the HIV type 1 coreceptor-binding site on gp120. *Proc. Natl Acad. Sci. USA* **101,** 2706–2711 (2004).

Key: Leader sequences (underlined), CD4 domains 1 and 2 (red), linker regions (black), antibody hinge and Fc regions (cyan), CCR5-mimetic sequences (green), and mutations (grey highlight).

CD4-Ig CD5 leader sequence (human); CD4 domains 1 and 2 (human); short AADP linker; IgG1 hinge and Fc (human).

MPMGSLQPLATLYLLGMLVASVLAKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYIC
EVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAAADPEPKSCDKTHT
CPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIE
KTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSL
SLSPGK

eCD4-Ig CD5 leader sequence (human); CD4 domains 1 and 2 (human); short AADP linker; IgG1 hinge and Fc (human); tetraglycine linker; CCR5mim1.

MPMGSLQPLATLYLLGMLVASVLAKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYIC
EVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAAADPEPKSCDKTHT
CPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIE
KTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSL
SLSPGKGGGGGDYADYDGGYYYDMD

Fusion 1 CD5 leader sequence (human); CCR5mim1; tetraglycine linker; CD4 domains 1 and 2 (human); short AADP linker; IgG1 hinge and Fc (human).

MPMGSLQPLATLYLLGMLVASVLAGDYADYDGGYYYDMDGGGGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQG
NFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKID
IVVLAAADPEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWL
NGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNV
FSCSVMHEALHNHYTQKSLSLSPGK

Fusion 2 CD5 leader sequence (human); CD4 domains 1 and 2 (human); short AA linker; tetraglycine linker; CCR5mim1; short DP linker; IgG1 hinge and Fc (human)

MPMGSLQPLATLYLLGMLVASVLAKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYIC
EVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAAAGGGGGDYADYDG
GYYYDMDDPEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWL
NGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNV
FSCSVMHEALHNHYTQKSLSLSPGK

eCD4-Ig^{mim2} CD5 leader sequence (human); CD4 domains 1 and 2 (human); short AADP linker; IgG1 hinge and Fc (human); tetraglycine linker; CCR5mim2 (Y4A difference from CCR5mim1 highlighted).

MPMGSLQPLATLYLLGMLVASVLAKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYIC
EVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAAADPEPKSCDKTHT
CPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIE
KTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSL
SLSPGKGGGGGDYYDYDGGYYYDMD

eCD4-Ig^{Q40A} CD5 leader sequence (human); CD4 domains 1 and 2 (human) with Q40A mutation (highlighted); short AADP linker; IgG1 hinge and Fc (human); tetraglycine linker; CCR5mim1.

MPMGSLQPLATLYLLGMLVASVLAKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNAGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYIC
EVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAAADPEPKSCDKTHT
CPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIE
KTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSL
SLSPGKGGGGGDYADYDGGYYYDMD

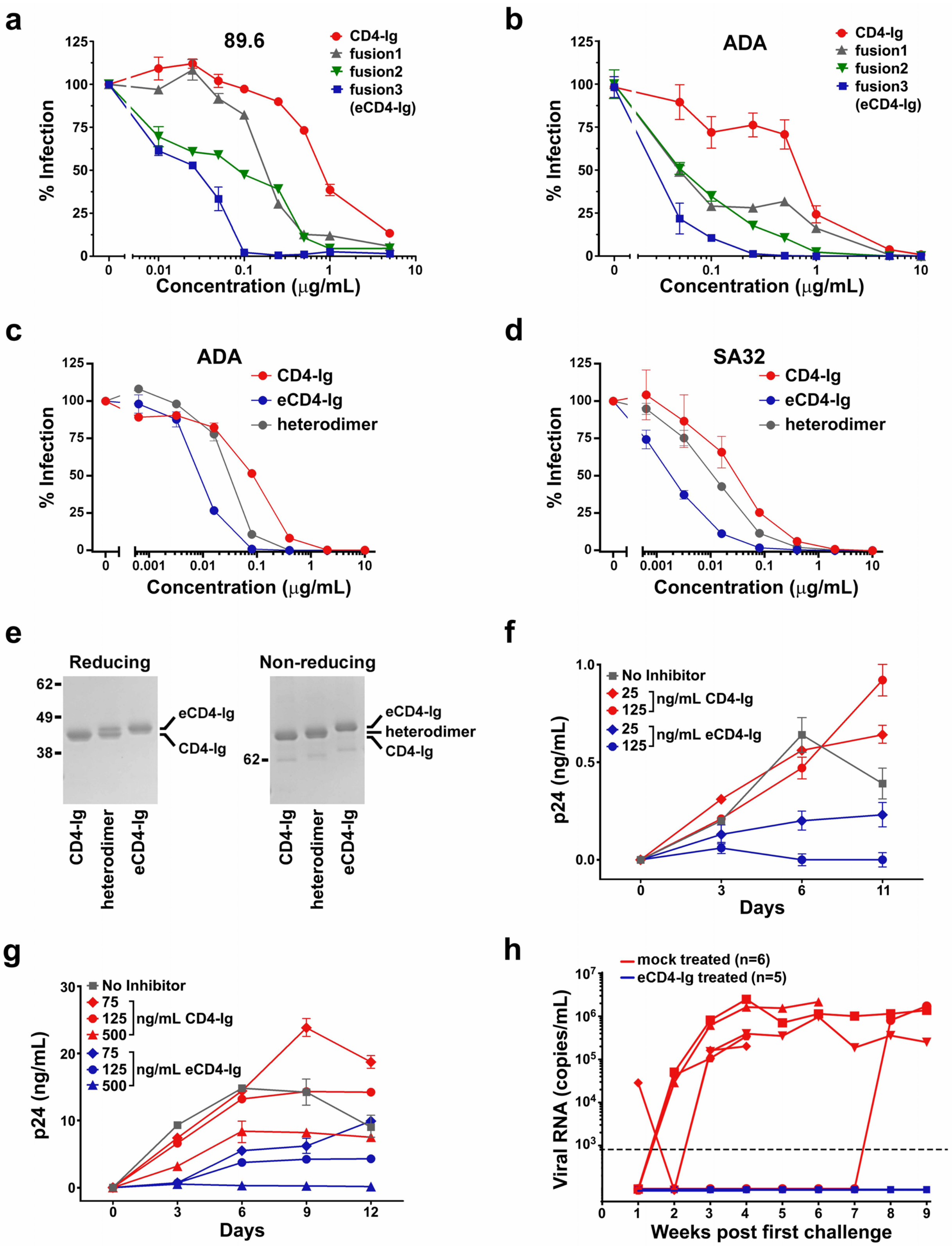
eCD4-Ig^{Q40A,mim2} CD5 leader sequence (human); CD4 domains 1 and 2 (human) with Q40A mutation (highlighted); short AADP linker; IgG1 hinge and Fc (human); tetraglycine linker; CCR5mim2 (difference from CCR5mim1 highlighted).

MPMGSLQPLATLYLLGMLVASVLAKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNAGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYIC
EVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAAADPEPKSCDKTHT
CPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIE
KTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSL
SLSPGKGGGGGDYYDYDGGYYYDMD

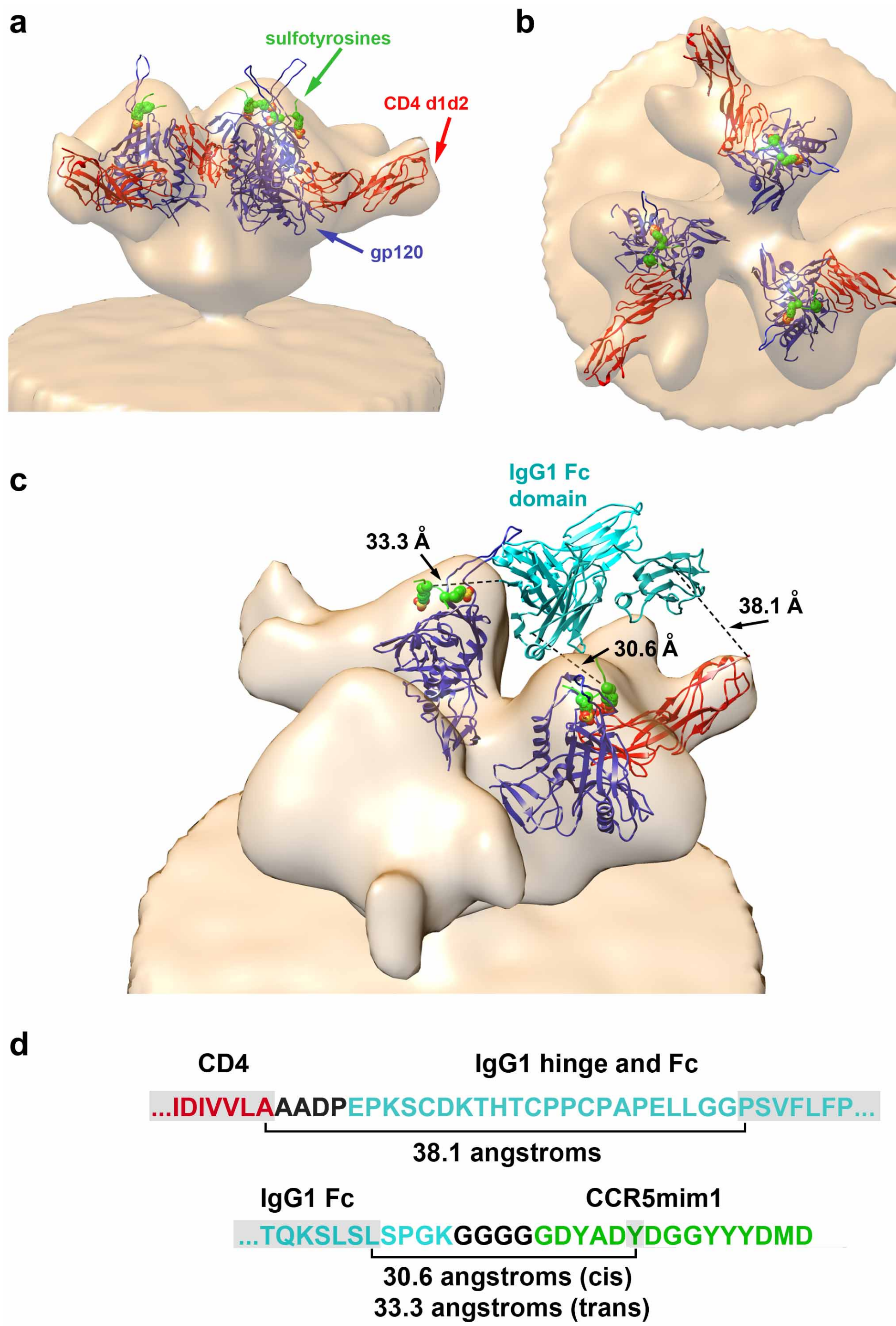
rh-eCD4-Ig (rh-eCD4-IgG2^{I39N,mim2}) CD4 leader sequence (rhesus); CD4 domains 1 and 2 (rhesus) with I39N mutation (highlighted); IgG2 hinge and Fc (rhesus); tetraglycine linker; CCR5mim2 (difference from CCR5mim1 highlighted).

MNRGIPFRHLLLVLQLALLPAVTQGKKVVLGKKGDTVELTCNASQKKNTQFHWKNSNQIKILGNQGSFLTKGPSKLSDRADSRKSLWDQGCFSMIIKNLKIEDSDTYI
CEVENKKEEVELLVFGLTANSDTHLLEGQSLTLTLESPPGSSPSVKCRSPGGKNIQGGRTISVPQLERQDSGTWTCTVSQDQKTVEFKIDIVVLAFQKASSTGLPCRS
TCPPCPAELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSQEEPDVKFNWYVDGVEVHNAQTKPREEQFNSTYRVVSVLTVTHQDWLNGKEYTCKVSNKALPAPRQ
KTVSKTKGQPREPQVYTLPPPREELTKNQVSLTCLVKGFYPSDIVVEWASNGQPENTYKTTPPVLDSDGSYFLYSKLTVDKSRWQQGNTFSCSVMHEALHNHYTQKSL
SVSPGKGGGGGDYYDYDGGYYYDMD

Extended Data Figure 1 | Sequences of CD4-Ig and eCD4-Ig variants. The amino-acid sequences of CD4-Ig, eCD4-Ig, fusion1, fusion2, eCD4-Ig^{mim2}, eCD4-Ig^{Q40A}, eCD4-Ig^{Q40A,mim2} and rh-eCD4-Ig (rh-eCD4-IgG2^{I39N,mim2}) are shown. Leader peptides are underlined, CD4 domains 1 and 2 are indicated in red, Fc domains are indicated in cyan, CCR5-mimetics peptides are indicated in green, and linker sequences are shown in black.



Extended Data Figure 2 | Additional characterization of eCD4-Ig. **a**, **b**, Experiments similar to those of Fig. 1b except that CD4-Ig (red), fusion1 (grey), fusion2 (green) and fusion3 (eCD4-Ig; blue) are compared using HIV-1 pseudotyped with the envelope glycoproteins of the 89.6 (**a**) or ADA (**b**) isolates. **c**, **d**, Experiments similar to those in Fig. 1e except that CD4-Ig (red), eCD4-Ig (blue) or heterodimers thereof (grey) are compared. **e**, CD4-Ig, eCD4-Ig and the CD4-Ig/eCD4-Ig heterodimer assayed in **c**, **d** and Fig. 1e were analysed by SDS–PAGE and stained with Coomassie blue under reducing (left) and non-reducing (right) conditions. **f**, **g**, Infectious 89.6 (**f**) or SG3 (**g**) HIV-1 was incubated with human PBMC in the presence of the indicated concentrations of CD4-Ig (red) or eCD4-Ig (blue), or without either inhibitor (grey). Culture supernatants were collected on the indicated day and viral p24 levels were measured by ELISA. **h**, Viral loads in RNA copies ml^{-1} are shown for each humanized mouse of Fig. 1f. Mice treated with eCD4-Ig are indicated with blue lines and mice treated with PBS are indicated with red lines. The 800 copies ml^{-1} limit of detection of this assay is indicated by a dashed line. Experiments in **a**–**g** were performed at least twice with similar results. Error bars denote s.e.m. of triplicates.



Extended Data Figure 3 | A model of eCD4-Ig bound to the HIV-1 Env trimer. **a**, The structure (2QAD) of gp120 (YU2 isolate) bound to the tyrosine-sulfated CD4i antibody 412d and CD4 domains 1 and 2 (ref. 10), was fitted into a cryoelectron micrograph of the HIV-1 envelope glycoprotein trimer (Env; Bal isolate) bound to CD4 (ref. 48). gp120 and CD4 domains 1 and 2 are shown in blue and red, respectively. 412d sulfotyrosines are represented as green (carbon), red (oxygen) and yellow (sulphur) spheres. The remainder of 412d was excluded for clarity. **b**, The same structure shown in **a** rotated 90° about the horizontal axis. Note that the sulfotyrosine-binding pockets are proximal to the trimer axis, whereas the C terminus of CD4 domain 2 is distal from the trimer axis, preventing both CD4 domains of CD4-Ig from simultaneously binding the same Env trimer. **c**, A model of how eCD4-Ig may associate with Env is presented. The Fc domain of human IgG1 (1FCC, cyan)[49] was positioned to be proximal to the gp120 sulfopeptide-binding pocket occupied by sulfotyrosine 100 (Tys 100) of the 412d heavy chain while avoiding steric interaction with Env. Tys 100 occupies a pocket in gp120 thought to bind CCR5 sulfotyrosine 10 (ref. 50). This pocket is also critical for binding of CCR5mim1 and CCR5mim 2 (refs 20, 22). In this model, the Fc domain is oriented to allow each eCD4-Ig sulfopeptide to engage a different gp120 protomer[24]. A single CD4 domain also binds one of the sulfopeptide-bound protomers. Distances between the C terminus of CD4 and the N terminus of one Fc domain monomer (38.1 Å), between the C terminus of the Fc domain and Tys 100 pocket of the CD4-bound gp120 protomer (30.6 Å), and between the C terminus of the Fc domain and Tys 100 pocket of an adjacent gp120 protomer (33.3 Å), are indicated. **d**, Residues not visible in the crystal structures used to construct this model are shown between brackets. In the model shown in **c**, these residues span the distances indicated. Note that these distances are well under the extension of a typical beta strand. CD4-, IgG1- and CCR5mim1-derived residues are shown in red, cyan, and green, respectively, with linker regions shown in black. Residues visible in the crystal structures, including the CCR5mim1 sulfotyrosine presumed to fill the Tys 100 pocket, are highlighted in grey. Modelling was performed using UCSF Chimera version 1.8.

a IC50 values

Virus	Clade	Tier	CD4-Ig	eCD4-Ig variants: eCD4-Ig	mim2	Q40A	Q40A, mim2	Fold	Geometric Mean: eCD4-Ig variants	Geometric Mean: CD4bs bNAbs
SF162.LS	B	1A	<0.001	<0.001	<0.001	<0.001	<0.001	1.0	0.001	0.048
BaL.26	B	1B	0.006	<0.001	<0.001	0.001	<0.001	>6	0.001	0.015
DJ263.8	AG	1B	0.018	0.001	0.001	0.004	0.001	12.7	0.001	0.032
ZM109F.PB4	C	1B	0.042	0.002	0.002	0.003	0.002	19.0	0.002	0.063
TV1.21	C	1B	0.045	0.002	0.001	0.008	0.001	22.5	0.002	ND
3365.v2.c20	A	2	0.066	0.002	0.002	0.016	0.002	19.6	0.003	0.020
SIVmac251.30	SIV	3	0.197	0.006	0.007	0.144	0.025	10.0	0.020	>50
QH0692.42	B	2	0.396	0.013	0.013	0.004	0.002	65.3	0.006	0.681
THRO4156.18	B	2	0.508	0.018	0.017	0.009	0.004	49.6	0.010	2.855
Q769.d22	A	2	0.661	0.021	0.028	0.021	0.018	30.4	0.022	0.011
3016.v5.c45	D	2	0.681	0.028	0.029	0.19	0.026	15.2	0.045	0.268
Q259.d2.17	A	2	2.141	0.076	0.079	0.081	0.048	30.8	0.070	0.034
T33-7	AG	3	3.512	0.053	0.053	0.005	0.003	245.1	0.014	0.023
T25118	AG	3	6.071	0.275	0.206	0.125	0.016	58.8	0.103	0.996
AC10.0.29	B	2	6.284	0.179	0.181	0.443	0.331	23.9	0.263	1.466
ZM135M.PL10a	C	2	6.373	0.42	0.272	0.094	0.042	43.7	0.146	0.289
PVO.4	B	3	9.506	0.212	0.214	0.032	0.025	122.5	0.078	0.162
CH115.12	B	3	25.676	0.786	0.554	0.586	0.255	50.8	0.505	ND
Du156.12	C	2	26.267	0.782	0.694	0.173	0.067	93.3	0.282	0.044
T257-31	AG	2	40.001	0.509	0.416	3.429	0.507	51.4	0.779	0.560
X1193_C1	G	2	40.218	0.367	0.476	0.111	0.028	263.5	0.153	ND
TRJO4551.58	B	3	>50	0.717	0.581	0.12	0.052	>221.4	0.226	0.052
TRO.11	B	2	>50	0.917	0.759	0.225	0.131	>132.1	0.378	0.227
R1166.c1	AE	2	>50	1.137	0.983	1.246	0.226	>66.8	0.749	0.890

b IC50 values

Virus	Clade	CD4-Ig	eCD4-Ig variants: mim2	Q40A, mim2	Fold	NIH 45-46	3BNC117	VRC01
TV1.29	C	0.055	0.001	<0.001	>55.0	>50	>50	>50
Du123.06	C	0.082	<0.001	<0.001	>82.0	>50	0.183	13.6
57128.vrc15	D	0.243	0.007	0.004	45.9	>50	0.432	>50
89-F1_2_25	CD	0.491	0.022	0.008	37.0	>50	>50	ND
CH070.1	BC	0.507	0.010	0.003	92.6	>50	7.89	18.7
CNE7	BC	0.576	0.028	0.005	48.7	0.014	>50	0.54
Du172.17	C	0.821	0.022	0.003	101.1	>50	1.19	>50
Du151.02	C	0.823	0.046	0.011	36.6	>50	>50	7.7
6545.v4.c1	AC	0.835	0.043	0.123	11.5	>50	>50	>50
CAP210.2.00.E8	C	1.033	0.029	0.010	60.7	>50	8.16	>50
ZM247v1(rev-)	C	1.079	0.043	0.056	22.0	2.185	>50	ND
242-14	AG	1.192	0.042	0.006	75.1	>50	>50	>50
X2088.c9	G	1.437	0.075	0.028	31.4	>50	>50	>50
Ce1172_H1	C	1.619	0.127	0.014	38.4	>50	>50	ND
1394C9G1 (rev-)	C	1.738	0.086	0.011	56.5	0.027	>50	ND
T278-50	AG	1.826	0.203	0.665	5.0	>50	>50	>50
CNE15	BC	1.848	0.112	0.006	71.3	0.005	>50	0.08
6540.v4.c1	AC	1.987	0.073	0.192	16.8	>50	>50	>50
6322.v4.c1	C	2.282	0.100	0.029	42.4	>50	>50	>50
6631.v3.c10	C	2.396	0.102	0.197	16.9	>50	>50	>50
7165.18	B	2.469	0.143	0.038	33.5	>50	6.54	>50
CNE20	BC	2.583	0.123	0.050	32.9	3.682	>50	ND
6471.v1.c16	C	2.662	0.164	0.036	34.6	>50	>50	>50
CH038.12	BC	3.490	0.139	0.176	22.3	0.059	>50	0.379
00836-2.5	C	3.566	0.093	0.025	74.0	<0.001	>50	0.128
A03349M1.vrc4a	D	3.874	0.176	0.018	68.8	>50	0.512	4.66
6545.v3.c13	AC	3.960	0.092	0.501	18.4	>50	>50	ND
Du422.1	C	4.268	0.157	0.036	56.8	>50	>50	>50
H086.8	B	6.522	0.071	0.088	82.5	>50	>50	>50
T251-18	AG	7.626	0.188	0.035	94.0	0.863	0.203	3.58
CAP45.2.00.G3	C	7.661	0.140	0.029	120.2	>50	0.589	9.47
T250-4	AG	8.961	0.218	0.006	247.8	>50	>50	>50
T266-60	AG	10.982	0.377	0.029	105.0	0.363	0.032	0.353
0077.V1.C16	C	12.703	0.245	0.042	125.2	0.160	>50	1.04
3718.v3.c11	A	15.392	0.676	0.401	29.6	2.049	>50	0.218
191955_A11	A	20.979	0.454	0.757	35.8	0.116	>50	ND
3988.25	B	21.728	0.623	1.332	23.9	0.071	>50	2.1
3637.v5.c3	C	28.616	0.655	0.046	164.9	0.799	>50	4.09
3817.v2.c59	CD	30.770	1.453	0.130	70.8	>50	0.216	>50
620345.c1	AE	46.395	0.515	0.051	286.3	>50	>50	>50

µg/ml	<0.01	0.01-0.1	0.1-1	1 to 10	10 to 50	>50	NOT DONE
nM	<0.1	0.1 to 1	1 to 10	10 to 100	100-500	>500	

Extended Data Figure 4 | IC$_{50}$ values of eCD4-Ig variants against neutralization-resistant isolates. **a**, The IC$_{50}$ values (µg ml^{-1}) of CD4-Ig, eCD4-Ig, eCD4-Ig^{mim2} (mim2), eCD4-Ig^{Q40A} (Q40A) and eCD4-Ig^{Q40A,mim2} (Q40A,mim2) against 24 HIV-1 and SIV isolates selected for their neutralization resistance are shown. The clade and tier of each isolate is listed. HIV-1 pseudotyped with the indicated envelope glycoprotein was incubated in triplicate with TZM-bl cells and varying concentrations of CD4-Ig or eCD4-Ig variant. Luciferase activity was determined 2 days after infection. 'Fold' indicates the ratio of the IC$_{50}$ value of CD4-Ig to the geometric mean of the IC$_{50}$ values of the assayed eCD4-Ig variants. The geometric mean of eCD4-Ig variants and the CD4bs antibodies 3BCN117, NIH45-46 and VRC01 calculated from values reported in previously[4,6] are shown in the two right-most columns. **b**, Neutralization studies similar to those in **a** except that the IC$_{50}$ values of CD4-Ig, eCD4-Ig^{mim2} (mim2), eCD4-Ig^{Q40A,mim2} (Q40A,mim2) and NIH45-46 were determined for a panel of 40 viral isolates selected for their resistance to the CD4bs bNAbs 3BNC117 and NIH45-46. IC$_{50}$ values of the CD4bs antibodies VRC01 and 3BNC117 listed in the two right-most columns were previously reported[4,6].

a IC80 values

Virus	Clade	Tier	CD4-Ig	eCD4-Ig variants: eCD4-Ig	eCD4-Ig variants: mim2	eCD4-Ig variants: Q40A	eCD4-Ig variants: Q40A, mim2	Fold	Geometric Mean: eCD4-Ig variants	Geometric Mean: CD4bs bNAbs	NIH45-46	3BNC117	VRC01
SF162.LS	B	1A	<0.001	<0.001	<0.001	<0.001	<0.001	1.0	<0.001	0.181	0.113	0.084	0.627
BaL.26	B	1B	0.041	0.002	0.002	0.002	0.001	24.4	0.002	0.061	0.052	0.028	0.154
DJ263.8	AG	1B	0.12	0.004	0.005	0.023	0.003	19.7	0.006	0.212	0.108	0.133	0.668
ZM109F.PB4	C	1B	0.316	0.014	0.014	0.045	0.025	14.6	0.022	0.329	0.373	0.242	0.394
TV1.21	C	1B	0.211	0.006	0.005	0.028	0.003	29.8	0.007	ND	ND	ND	ND
3365.v2.c20	A	2	0.294	0.016	0.015	0.101	0.012	12.7	0.023	0.090	0.185	0.044	ND
SIVmac251.30	SIV	3	1.443	0.191	0.148	5.605	0.841	2.4	0.604	>50	>50	>50	>50
QH0692.42	B	2	2.756	0.102	0.066	0.017	0.011	82.3	0.033	2.463	3.39	1.47	3
THRO4156.18	B	2	1.858	0.068	0.065	0.025	0.019	48.8	0.038	12.569	9.67	13.6	15.1
Q769.d22	A	2	7.248	0.119	0.155	0.116	0.097	60.4	0.120	0.067	0.09	0.045	0.074
3016.v5.c45	D	2	3.496	0.108	0.108	0.989	0.139	17.5	0.200	1.069	>50	3.35	0.341
Q259.d2.17	A	2	11.764	0.373	0.283	0.367	0.245	37.7	0.312	0.145	0.181	0.067	0.252
T33-7	AG	3	22.853	0.282	0.204	0.027	0.016	323.7	0.071	0.057	0.392	0.021	0.023
T25118	AG	3	40.544	1.259	0.743	0.578	0.064	94.0	0.431	4.783	12.5	0.858	10.2
AC10.0.29	B	2	35.017	0.693	0.535	1.691	1.608	34.9	1.002	2.916	2.22	>50	3.83
ZM135M.PL10a	C	2	45.557	1.677	0.83	0.526	0.433	60.7	0.750	1.708	2.89	0.28	6.16
PVO.4	B	3	>50	1.491	0.982	0.152	0.145	>118.0	0.424	0.517	0.474	0.294	0.99
CH115.12	B	3	>50	2.798	1.591	2.113	0.792	>30.3	1.652	ND	ND	ND	ND
Du156.12	C	2	>50	3.873	3.391	0.858	0.513	>32.2	1.551	0.127	0.07	0.121	0.244
T257-31	AG	2	>50	1.404	1.161	10.363	1.817	>21.2	2.354	2.047	2.13	0.694	5.8
X1193_C1	G	2	>50	2.364	2.092	0.66	0.119	>63.3	0.789	ND	ND	ND	ND
TRJO4551.58	B	3	>50	2.518	1.658	0.346	0.251	>64.4	0.776	0.179	0.111	0.216	0.239
TRO.11	B	2	>50	4.494	2.745	1.609	0.665	>26.2	1.906	0.955	6.4	0.125	1.09
R1166.c1	AE	2	>50	6.176	3.843	6.36	0.869	>14.8	3.384	3.415	7.02	0.805	7.05

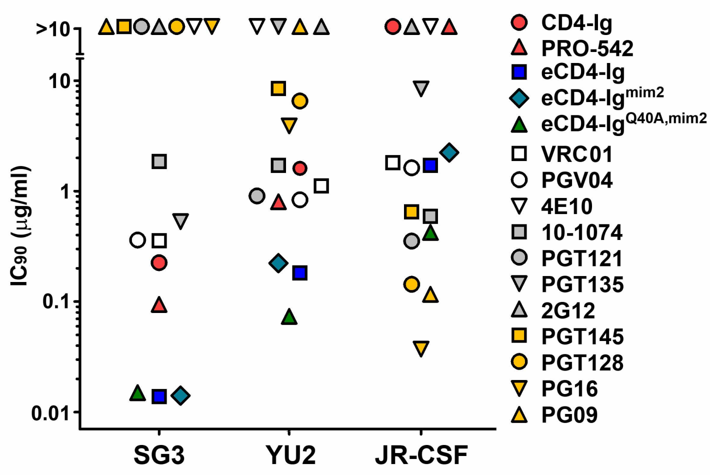
b IC80 values

Virus	Clade	CD4-Ig	eCD4-Ig variants: mim2	eCD4-Ig variants: Q40A, mim2	Fold	NIH 45-46	3BNC117	VRC01
TV1.29	C	0.196	0.005	0.003	50.6	>50	>50	>50
Du123.06	C	0.437	0.016	0.004	54.6	>50	1.17	>50
57128.vrc15	D	1.057	0.045	0.013	43.7	>50	1.84	>50
89-F1_2_25	CD	2.077	0.056	0.018	65.4	>50	>50	ND
CH070.1	BC	1.811	0.039	0.013	80.4	>50	50	50
CNE7	BC	2.861	0.111	0.017	65.9	0.085	>50	1.36
Du172.17	C	5.645	0.151	0.020	102.7	>50	2.46	>50
Du151.02	C	3.891	0.168	0.028	56.7	>50	>50	>50
6545.v4.c1	AC	5.089	0.182	0.483	17.2	>50	>50	>50
CAP210.2.00.E8	C	7.958	0.180	0.041	92.6	>50	>50	>50
ZM247v1(rev-)	C	5.702	0.189	0.245	26.5	>50	>50	ND
242-14	AG	18.026	0.378	0.052	128.6	>50	>50	>50
X2088.c9	G	5.026	0.252	0.082	35.0	>50	>50	>50
Ce1172_H1	C	13.312	0.460	0.062	78.8	>50	>50	ND
1394C9G1 (rev-)	C	39.283	0.732	0.082	160.3	0.161	>50	ND
T278-50	AG	14.186	0.971	3.419	7.8	>50	>50	>50
CNE15	BC	22.835	0.461	0.023	221.8	0.018	>50	0.28
6540.v4.c1	AC	17.253	0.337	1.338	25.7	>50	>50	>50
6322.v4.c1	C	10.366	0.425	0.121	45.7	>50	>50	>50
6631.v3.c10	C	7.926	0.233	0.666	20.1	>50	>50	>50
7165.18	B	12.997	0.558	0.172	42.0	>50	35.7	>50
CNE20	BC	23.876	0.719	0.346	47.9	>50	>50	ND
6471.v1.c16	C	13.816	0.623	0.137	47.3	>50	>50	>50
CH038.12	BC	19.888	0.527	0.937	28.3	0.221	>50	1.53
00836-2.5	C	38.764	0.744	0.138	121.0	0.007	>50	0.52
A03349M1.vrc4a	D	17.546	0.468	0.054	110.4	>50	2.34	28.1
6545.v3.c13	AC	15.344	0.442	1.748	17.5	>50	>50	ND
Du422.1	C	27.666	1.463	0.161	57.0	>50	>50	>50
H086.8	B	26.707	0.491	1.153	35.5	>50	>50	>50
T251-18	AG	41.025	0.509	0.102	180.0	3.134	0.858	10.2
CAP45.2.00.G3	C	41.712	0.714	0.120	142.5	>50	32.1	>50
T250-4	AG	33.879	0.769	0.032	216.0	>50	>50	>50
T266-60	AG	>50	1.503	0.159	>102.3	2.379	0.119	1.35
0077.V1.C16	C	>50	1.816	0.235	>76.5	0.767	>50	3.65
3718.v3.c11	A	>50	1.841	0.952	>37.8	>50	>50	4.99
191955_A11	A	>50	1.556	2.630	>24.7	0.592	>50	ND
3988.25	B	>50	2.288	11.626	>9.7	0.353	>50	>50
3637.v5.c3	C	>50	1.509	0.128	>113.8	3.069	>50	11.0
3817.v2.c59	CD	>50	5.133	0.366	>36.5	>50	0.752	>50
620345.c1	AE	>50	2.876	0.201	>65.8	>50	>50	>50

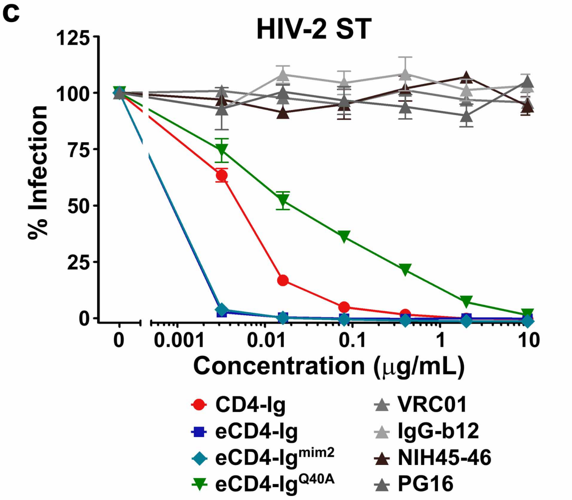
μg/ml	<0.01	0.01-0.1	0.1-1	1 to 10	10 to 50	>50	NOT DONE
nM	<0.1	0.1 to 1	1 to 10	10 to 100	100-500	>500	

Extended Data Figure 5 | IC_{80} values of eCD4-Ig variants against neutralization-resistant isolates. **a**, **b**, The IC_{80} values (μg ml^{-1}) of the experiments described in Extended Data Fig. 5a (**a**) and Extended Data Fig. 5b (**b**) are shown.

a



c



b

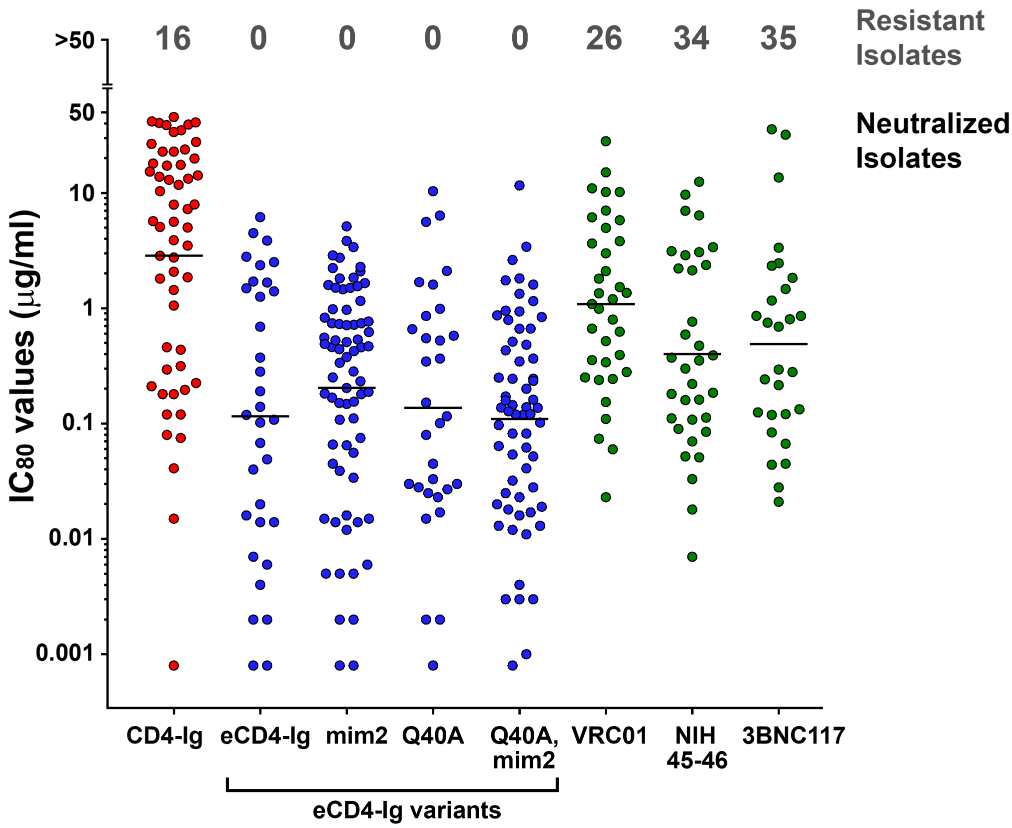
IC$_{50}$ values

Virus	CD4-Ig	eCD4-Ig variants: eCD4-Ig	eCD4-Ig variants: mim2	eCD4-Ig variants: Q40A, mim2	VRC01	VRC-PG04	PGT145	PG09	PG16	4E10	2G12	PGT121	PGT128	PGT135	10-1074	PRO542
SG3	0.027	0.002	0.002	0.003	0.060	0.060	> 10	> 10	> 10	1.540	0.981	0.121	> 10	0.076	0.154	0.012
+/-	0.002	0.000	0.000	0.001	0.009	0.014	0.000	0.000	0.000	0.499	0.153	0.023	0.000	0.020	0.021	0.002
YU2	0.293	0.030	0.035	0.015	0.134	0.116	1.842	3.702	0.145	> 10	> 10	0.121	0.244	> 10	0.280	0.115
+/-	0.043	0.003	0.004	0.000	0.005	0.002	0.279	0.481	0.009	0.000	0.000	0.021	0.059	0.000	0.028	0.014
JR-CSF	3.175	0.237	0.270	0.064	0.285	0.188	0.052	0.013	0.004	3.103	0.995	0.055	0.020	0.859	0.097	1.588
+/-	0.635	0.023	0.043	0.006	0.074	0.061	0.004	0.002	0.000	0.120	0.096	0.020	0.008	0.066	0.014	0.115

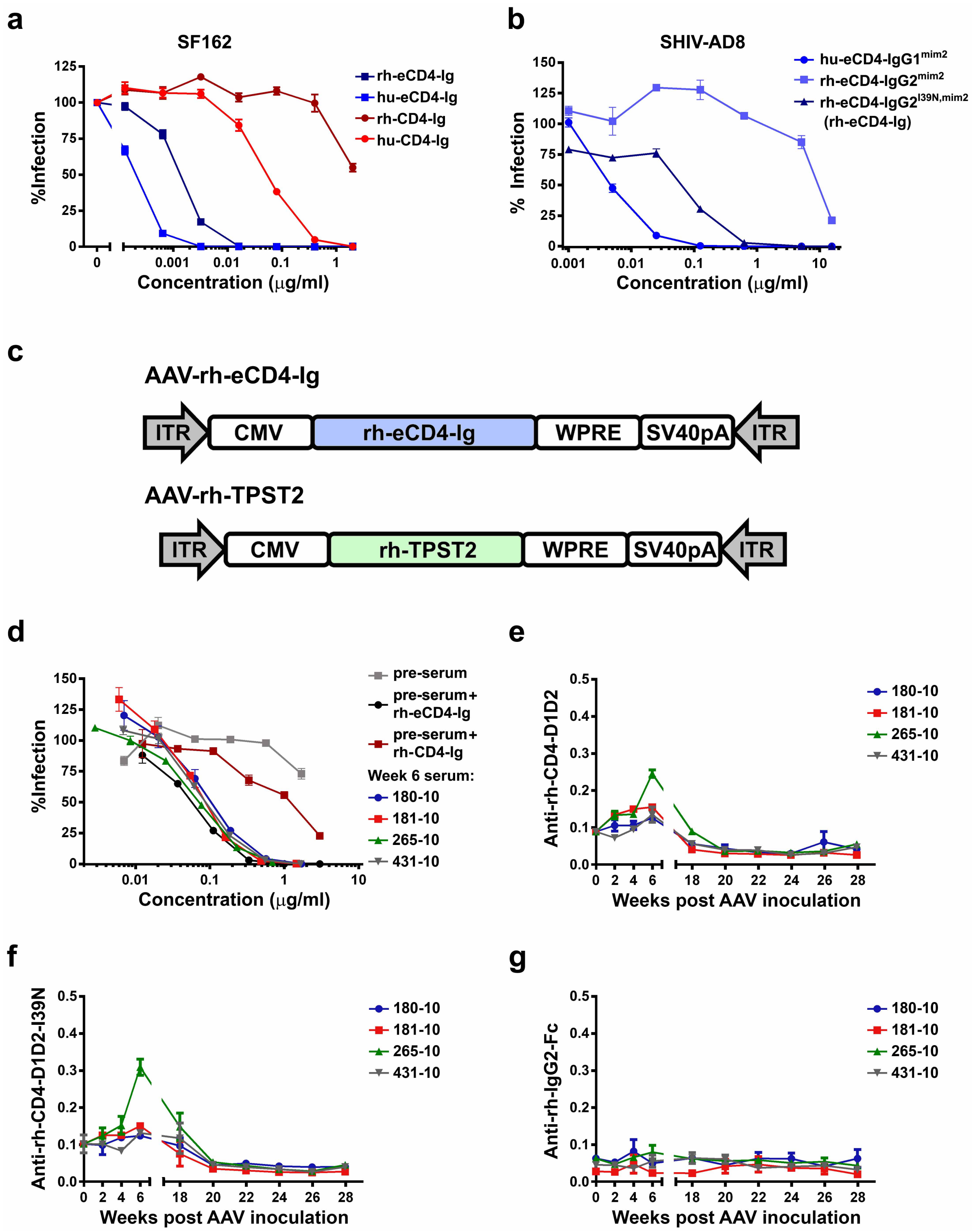
IC$_{90}$ values

Virus	CD4-Ig	eCD4-Ig variants: eCD4-Ig	eCD4-Ig variants: mim2	eCD4-Ig variants: Q40A, mim2	VRC01	VRC-PG04	PGT145	PG09	PG16	4E10	2G12	PGT121	PGT128	PGT135	10-1074	PRO542
SG3	0.225	0.014	0.014	0.015	0.356	0.360	> 10	> 10	> 10	> 10	> 10	> 10	> 10	0.529	1.860	0.094
+/-	0.023	0.001	0.000	0.001	0.018	0.031	> 10	0.000	0.000	0.000	0.000	0.000	0.000	0.178	0.012	0.013
YU2	1.607	0.182	0.223	0.074	1.113	0.836	8.519	> 10	3.896	> 10	> 10	0.906	6.544	> 10	1.717	0.802
+/-	0.040	0.013	0.017	0.001	0.012	0.030	0.992	0.000	1.341	0.000	0.000	0.157	1.125	0.000	0.074	0.096
JR-CSF	> 10	1.709	2.237	0.423	1.812	1.629	0.651	0.117	0.037	> 10	> 10	0.353	0.144	8.429	0.593	> 10
+/-	0.000	0.087	0.401	0.077	0.134	0.255	0.303	0.006	0.005	0.000	0.000	0.038	0.042	0.137	0.096	0.000

Extended Data Figure 6 | Further comparison of eCD4-Ig and HIV-1 neutralizing antibodies. **a**, IC$_{90}$ values for the same experiments shown in Fig. 2a, presented in the same format. **b**, Numeric IC$_{50}$ and IC$_{90}$ values of the experiment shown in **a** and Fig. 2a are shown, using the same colour coding of Extended Data Figs 4 and 5. The s.e.m. of triplicates are indicated below their IC$_{50}$ and IC$_{90}$ values. **c**, Experiments similar to those in Fig. 2b except that HIV-1 pseudotyped with the Env of the HIV-2 isolate ST was incubated with the indicated concentrations of CD4-Ig, eCD4-Ig variants or the CD4bs antibodies IgG-b12, VRC01 or NIH45-46. Error bars denote s.e.m. of triplicates.



Extended Data Figure 7 | Summary of IC_{80} values for HIV-1, HIV-2 and SIV neutralization studies. The IC_{80} values from studies of Figs 1b, 2a, b, and Extended Data Figs 4–6 are plotted. The number of isolates resistant to 50 μg ml^{-1} of the indicated inhibitors are indicated at the top. Geometric means are calculated for neutralized isolates and indicated with horizontal lines.



Extended Data Figure 8 | Additional characterization of rh-eCD4-Ig. **a**, An experiment similar to that in Fig. 2b, except that rhesus and human CD4-Ig and eCD4-Ig are compared for their ability to neutralize HIV-1 pseudotyped with the SF162 envelope glycoprotein. All variants have wild-type rhesus or human CD4 domains. Note that variants bearing rhesus CD4 are markedly less potent at neutralizing HIV-1. **b**, Experiment similar to **a** and Fig. 2b except that human eCD4-Ig^{mim2} and its rhesus analogue bearing or not bearing the Ile39Asn mutation are compared using SHIV-AD8. Note that the Ile39Asn mutation largely restores the neutralization activity of rhesus eCD4-Ig^{mim2}. **c**, A representation of the AAV vectors used in the non-human primate studies of Fig. 4. Rh-eCD4-Ig (rh-eCD4-IgG2^{I39N,mim2}; blue) and rhesus tyrosine protein sulfotransferase 2 (TPST2; green) were introduced into a single-stranded AAV vector downstream of a CMV promoter. A woodchuck response element (WPRE), used to promote expression, and the SV40 polyadenylation signal (SV40pA) were also included. AAV inverted terminal repeats (ITR) are indicated in grey arrows. **d**, An experiment similar to that in Fig. 4d except that sera from week 6 were analysed. **e**–**g**, Experiments similar to those in Fig. 4f–h except that the reactivity of rhesus sera was examined for a construct bearing wild-type rhesus CD4 domains 1 and 2 fused to the human IgG1 Fc domain (**e**), one bearing rhesus CD4 domains 1 and 2 with the Ile39Asn mutation, again fused to the human IgG1 Fc domain (**f**), or the antibody NIH45-46 fused to the rhesus IgG2 constant regions, used here to present the rhesus IgG2 Fc domain (**g**). Experiments in **a**, **b** and **d**–**g** represent at least two with similar results. Error bars denote s.e.m. of triplicates.

Extended Data Table 1 | Potencies and breadth of well-characterized broadly neutralizing antibodies

Antibody	IC_{50}	IC_{80}	IC_{50} < 50 μg/ml	IC_{80} < 5 μg/ml
10-1074	0.053	0.217	57.8%	51.2%
35O22	0.057	n.a.	61.9%	n.a. (<61.9%)
PGT121	0.060	0.274	63.0%	47.9%
PGT128	0.069	n.a.	62.9%	n.a. (<62.9%)
PG16	0.092	0.178	55.6%	43.6%
3BNC117	0.111	0.345	82.2%	61.0%
VRC07	0.114	0.187	83.2%	86.0%
NIH45-46	0.139	0.540	83.7%	57.4%
12A12	0.171	1.101	93.2%	68.9%
PG9	0.176	0.427	77.3%	62.0%
10E8	0.262	1.536	98.3%	75.5%
VRC01	0.306	0.913	88.0%	69.7%

A summary of antibody neutralization potencies compiled using the Los Alamos National Laboratory Database CATNAP tool (http://www.hiv.lanl.gov/components/sequence/HIV/neutralization/main.comp). The geometric mean IC_{50} and IC_{80} values are listed for the indicated bNAbs against all reported isolates, excluding those with values greater than 50 μg ml^{-1}. The percentage of isolates neutralized with IC_{50} values less than 50 μg ml^{-1}, or with IC_{80} values less than 5 μg ml^{-1} are shown. bNAbs are ranked by their geometric mean IC_{50} values. See Fig. 3 and Extended Data Fig. 7 for comparisons of eCD4-Ig variants with the bNAbs NIH45-46, 3BNC117 and VRC01.

doi:10.1038/nature14232

Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome

Benoit Chassaing[1], Omry Koren[2], Julia K. Goodrich[3], Angela C. Poole[3], Shanthi Srinivasan[4], Ruth E. Ley[3] & Andrew T. Gewirtz[1]

The intestinal tract is inhabited by a large and diverse community of microbes collectively referred to as the gut microbiota. While the gut microbiota provides important benefits to its host, especially in metabolism and immune development, disturbance of the microbiota–host relationship is associated with numerous chronic inflammatory diseases, including inflammatory bowel disease and the group of obesity-associated diseases collectively referred to as metabolic syndrome. A primary means by which the intestine is protected from its microbiota is via multi-layered mucus structures that cover the intestinal surface, thereby allowing the vast majority of gut bacteria to be kept at a safe distance from epithelial cells that line the intestine[1]. Thus, agents that disrupt mucus–bacterial interactions might have the potential to promote diseases associated with gut inflammation. Consequently, it has been hypothesized that emulsifiers, detergent-like molecules that are a ubiquitous component of processed foods and that can increase bacterial translocation across epithelia *in vitro*[2], might be promoting the increase in inflammatory bowel disease observed since the mid-twentieth century[3]. Here we report that, in mice, relatively low concentrations of two commonly used emulsifiers, namely carboxymethylcellulose and polysorbate-80, induced low-grade inflammation and obesity/metabolic syndrome in wild-type hosts and promoted robust colitis in mice predisposed to this disorder. Emulsifier-induced metabolic syndrome was associated with microbiota encroachment, altered species composition and increased pro-inflammatory potential. Use of germ-free mice and faecal transplants indicated that such changes in microbiota were necessary and sufficient for both low-grade inflammation and metabolic syndrome. These results support the emerging concept that perturbed host–microbiota interactions resulting in low-grade inflammation can promote adiposity and its associated metabolic effects. Moreover, they suggest that the broad use of emulsifying agents might be contributing to an increased societal incidence of obesity/metabolic syndrome and other chronic inflammatory diseases.

Mice were administered the emulsifiers carboxymethylcellulose (CMC) or polysorbate-80 (P80) via drinking water (1.0% w/v or v/v, respectively) for 12 weeks. P80 has been studied for toxicity and carcinogenic potential[2,4,5] and is approved by the US Food and Drug Administration for use in select foods at up to 1.0%. CMC has not been extensively studied but is deemed ‘generally regarded as safe (GRAS)’ and used in various foods at up to 2.0%[3,6]. We used wild-type C57Bl/6 mice and two engineered strains of mice, namely *Il10*$^{-/-}$ and *Tlr5*$^{-/-}$, that are prone to developing shifts in microbiota composition and inflammation[7,8]. While observations that loss of Toll-like receptor (TLR) function alters microbiota composition and promotes inflammation were suggested to be an artefact of mouse husbandry practices[9], the fact that epithelial-cell-specific deletion of TLR5 altered microbiota composition relative to wild-type siblings and resulted in low-grade inflammation/metabolic syndrome indicates that, at least for TLR5, this is not the case[10].

Confocal microscopy, using mucus-preserving Carnoy fixation[1,11], indicated that the closest bacteria resided, on average, about 25 μm from epithelial cells with no bacteria observed within 10 μm (Fig. 1a, d). In contrast, emulsifier-treated mice exhibited some bacteria in apparent contact with the epithelium while the average distance was reduced by more than 50% (Fig. 1b–d). Such microbiota encroachment correlated with reduced mucus thickness. This altered mucus thickness was not correlated with altered expression of MUC2 (Extended Data Fig. 1a–d). *Il10*$^{-/-}$ and *Tlr5*$^{-/-}$ mice exhibited basal microbiota encroachment that was enhanced by both CMC and P80 (Fig. 1e–h and Extended Data Fig. 1e–h). Emulsifiers did not affect total levels of faecal bacteria in wild-type, *Il10*$^{-/-}$ or *Tlr5*$^{-/-}$ mice but resulted in more than a twofold increase in bacteria adherent to the colons of wild-type and *Il10*$^{-/-}$ mice (Fig. 1i, k and Extended Data Fig. 1i–l). Moreover, interrogation of microbiota composition with 16S RNA sequencing, followed by phylogenic analysis, and use of the unweighted UniFrac algorithm to compare community structure[12] revealed that both CMC and P80 dramatically altered microbiota composition in both faecal and intestinal-adherent bacteria of wild-type, *Il10*$^{-/-}$ and *Tlr5*$^{-/-}$ mice (Fig. 1j, l and Extended Data Fig. 2a–o and Supplementary Tables 1 and 2).

In order to verify that differences in microbiota composition were a consequence of emulsifier treatment rather than simply reflecting that mice shared cages during treatment, which results in microbiota composition clustering in coprophagic animals, multiple litters were split equally into groups that were to receive water, CMC or P80 (three separate cages per condition as shown in Extended Data Fig. 3). As expected, this experimental design avoided pre-treatment clustering. However, microbiota composition clustered strongly following treatment irrespective of cage clustering in both male and female mice with wild-type and *Tlr5*$^{-/-}$ genotypes (Extended Data Fig. 3), confirming that emulsifiers alter microbiota composition.

Alterations in microbiota composition included reduced levels of Bacteroidales, associated with health[13–15], and increased levels of several mucolytic operational taxanomic units (OTUs) including *Ruminococcus gnavus*[16] (Extended Data Figs 2p–t and 4 and Supplementary Tables 1 and 2). In *Il10*$^{-/-}$ mice, both CMC and P80 induced a marked reduction in microbial diversity, bloom in Verrucomicrobia phyla (especially *Akkermansia muciniphila*; Extended Data Fig. 2p–t and Supplementary Tables 1 and 2)[17], and enriched mucosa-associated inflammation-promoting Proteobacteria[18,19]. Emulsifier-induced changes in microbiota composition were sufficiently uniform such that a low number of OTUs (<15) afforded a reliable prediction of emulsifier treatment (Extended Data Fig. 4 and Supplementary Tables 1 and 2). Such distinguishing OTUs were spread across the Bacteria domain and several were common to all three host genotypes.

While most experiments herein used young mice (4 weeks old at the start of the experiment) based on the notion that microbiota are more prone to disturbance at an early age[18,20], administration of emulsifiers beginning at 4 months of age also destabilized and altered microbiota composition (Extended Data Fig. 5h–p). Specifically, both CMC and P80 resulted in decreased alpha diversity and reduced stability, as evidenced by an increased extent of week-to-week changes in principal

[1]Center for Inflammation, Immunity and Infection, Institute for Biomedical Sciences, Georgia State University, Atlanta, Georgia 30303, USA. [2]Faculty of Medicine, Bar Ilan University, Safed, 13115, Israel. [3]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA. [4]Digestive Diseases Division, Department of Medicine, Emory University School of Medicine, Atlanta, Georgia 30322, USA.

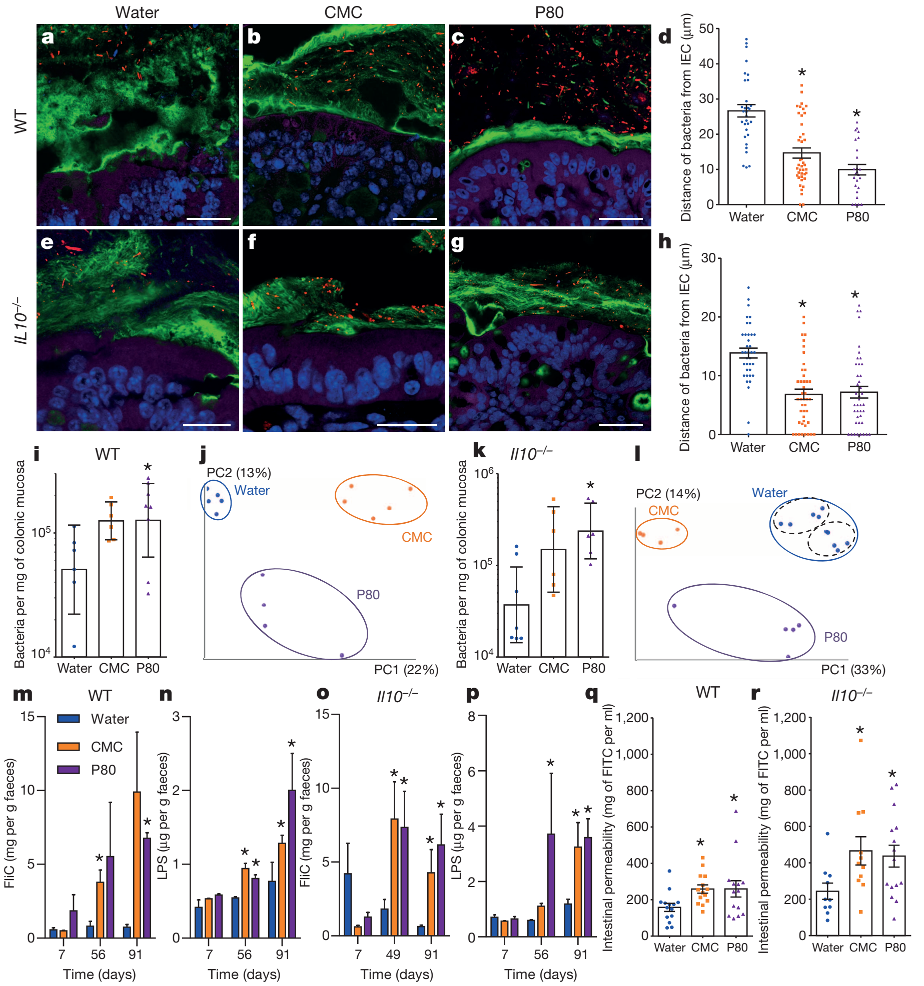


Figure 1 | Dietary emulsifiers alter microbiota localization, composition, and pro-inflammatory potential. **a-c**, **e–g**, Wild-type (WT; **a–c**) and *Il10*$^{-/-}$ (**e–g**) mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Confocal microscopy analysis of microbiota localization: MUC2, green; actin, purple; bacteria, red; and DNA, blue. Scale bar, 20 μm. Pictures are representative of 20 biological replicates. **d**, **h**, Distances of closest bacteria to intestinal epithelial cells (IEC) per condition over five high-powered fields per mouse. **i**, **k**, PCR-based quantification of bacterial load adhered to colonic mucosa. **j**, **l**, Principal coordinates analysis of the UniFrac distance matrix of wild-type (**j**) and *Il10*$^{-/-}$ (**l**) mice. Black dashed ellipses indicate mice sharing a cage during treatment. **m–p**, Bioactive levels of faecal flagellin (FliC) and LPS assayed with TLR5 and TLR4 reporter cells. **q–r**, Intestinal permeability measured by levels of serum FITC–dextran (4 kDa) following oral gavage. Data are the means ±s.e.m. or geometric means with 95% confidence interval (for **i** and **k**) ($n = 20$ for **a–h** and **m–p**; $n = 6$ for **i** except $n = 8$ for the P80-treated group; $n = 5$ for **j** except $n = 4$ for P80-treated group; $n = 6$ for **k** except $n = 7$ for water-treated group; for **l**, $n = 10$, 4 and 5 for water-, CMC- and P80-treated groups, respectively; $n = 14$ for **q**; for **r**, $n = 10$, 11 and 15 for water-, CMC- and P80-treated groups, respectively). Points in **i–l** and **q–r** are from individual mice. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test or two-way ANOVA corrected for multiple comparisons with a Bonferroni test; $*P < 0.05$ compared to water-treated group.

coordinates, and a greater extent of change over the course of the experiment that outweighed effects of cage clustering (Extended Data Fig. 5h–p). Thus, emulsifiers can alter the gut microbiota composition of hosts of a broad age range.

Microbiota composition influences the ability of a microbiota to activate innate immune signalling[21]. Thus, we measured the capacity of faeces from control and emulsifier-treated mice to activate pro-inflammatory gene expression via the lipopolysaccharide (LPS) and flagellin receptors TLR4 and TLR5, respectively. Exposure to emulsifiers increased faecal levels of bioactive LPS and flagellin in wild-type, *Il10*$^{-/-}$ and *Tlr5*$^{-/-}$ mice (Fig. 1m–p and Extended Data Fig. 5q–r). Emulsifier treatment also increased gut permeability in wild-type and *Il10*$^{-/-}$ mice (Fig. 1q, r), which correlated with increased levels of serum antibodies to flagellin and LPS (Extended Data Fig. 5s–v), thought to reflect gut permeability[22]. Thus, chronic exposure to dietary emulsifiers results in erosion of the protective function of the mucus, increased bacterial adherence and a more pro-inflammatory microbiota.

The hallmark of active colitis is the presence of immune cell infiltrates, which is typically paralleled by changes in gross colon morphology and elevated levels of the leukocyte enzyme myeloperoxidase and pro-inflammatory markers. Based on such criteria, emulsifiers promoted the extent and incidence of colitis in both *Il10*$^{-/-}$ and *Tlr5*$^{-/-}$ mice (Fig. 2a–e and Extended Data Figs 6a, b, h and 7a–g). Emulsifiers did not induce overt colitis in wild-type mice but did result in subtle histopathologic and gross evidence of chronic intestinal inflammation, including epithelial damage and shortened colons (Fig. 2f–i and Extended Data Fig. 6c–g). Faecal lipocalin 2 (LCN2) is a sensitive and broadly dynamic marker of intestinal inflammation in mice[23]. Emulsifier-treated wild-type mice exhibited modestly elevated faecal LCN2 levels 4 weeks after initial exposure (Fig. 2f). In *Il10*$^{-/-}$ and *Tlr5*$^{-/-}$ mice, basally elevated faecal LCN2 levels were further elevated (approximately tenfold) by 12 weeks of exposure to CMC and P80 (Fig. 2a and Extended Data Fig. 7a). Such robust colitis in *Il10*$^{-/-}$, but not *Tlr5*$^{-/-}$, correlated with *Bilophila* and *Helicobacter* enrichment, analogous to previous observations in *Il10*$^{-/-}$ mice (Supplementary Tables 2 and 3)[24,25]. The extent of inflammation inversely correlated with bacterial–epithelial distance in both wild-type and *IL10*$^{-/-}$ mice (Fig. 2j and Extended Data Fig. 6i–k). Thus, emulsifiers may promote robust colitis in susceptible hosts and induce low-grade inflammation in wild-type hosts.

Metabolic syndrome is associated with, and may be promoted by, low-grade inflammation[26]. Thus, we examined whether emulsifier-induced microbial dysbiosis and low-grade inflammation might promote this disorder in wild-type mice. Both CMC and P80 resulted in modest but significant gains in overall weight and a marked increase in adiposity as measured by fat mass (Fig. 3a, b). Such increased adiposity was associated with increased food consumption (Fig. 3c) that probably drove the increased adiposity. Emulsifier treatment also impaired glycaemic control as assessed by fasting blood glucose concentration (Fig. 3d) and glucose/insulin tolerance testing (Extended Data Fig. 7h, i).

Human exposure to dietary emulsifiers occurs in liquid and solid foods. Hence, we next examined whether incorporating emulsifiers into mouse chow would drive similar phenotypes. Supplementation of chow with 1.0% CMC or P80 fully mimicked the pro-inflammatory effects and metabolic changes induced by emulsifiers in drinking water (Fig. 3e, f and Extended Data Fig. 7l–t). While human emulsifier consumption is not a widely tracked parameter, their pervasive use in many foods exceeds the 1.0% level (in food or water) used herein[6]. Nonetheless, we next sought to define the minimum dose of emulsifiers that would produce

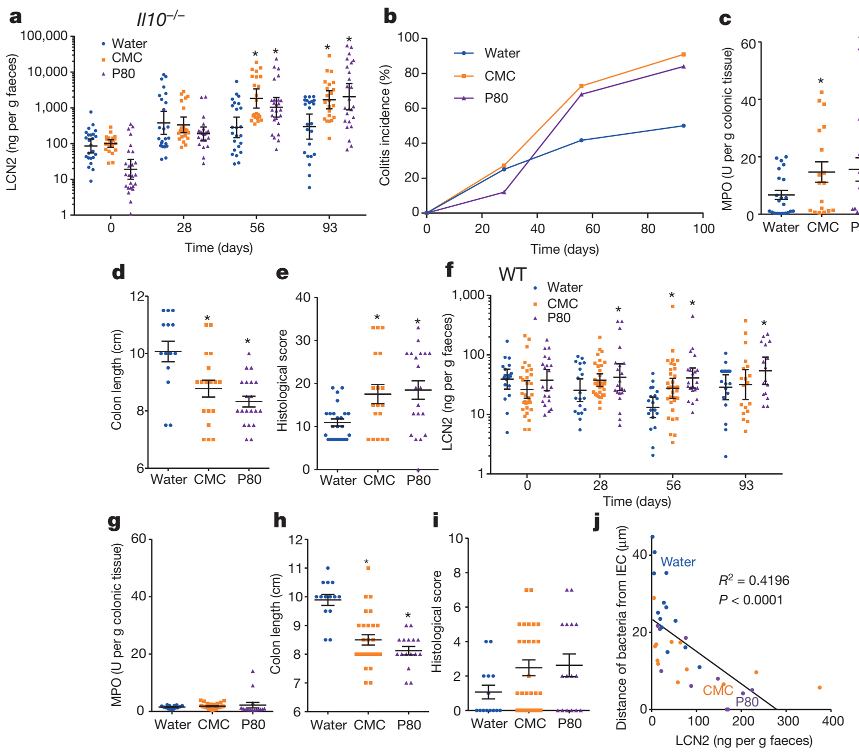


Figure 2 | Dietary emulsifiers promote colitis in susceptible mice and low-grade intestinal inflammation in wild-type mice. **a**–**i**, *Il10*$^{-/-}$ (**a**–**e**) and wild-type (WT; **f**–**i**) mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **a**, **f**, Faecal levels of the inflammatory marker LCN2 over time; **b**, colitis incidence (%) over time; **c**, **g**, myeloperoxidase (MPO) levels; **d**, **h**, colon lengths; and **e**, **i**, histological score after 12 weeks of exposure. **j**, Faecal levels of the inflammatory marker LCN2 plotted versus microbiota–epithelial distance obtained in Fig. 1d. IEC, intestinal epithelial cells. Data are the means ±s.e.m. or geometric means with 95% confidence intervals (the latter for **a** and **f**) ($n = 20$). Points are from individual mice. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test or two-way ANOVA corrected for multiple comparisons with a Bonferroni test; *$P < 0.05$ compared to water-treated group.

evidence of low-grade inflammation/metabolic syndrome. As little as 0.1% CMC resulted in modest increases in body weight and fasting glucose, while 0.5% resulted in clear evidence of low-grade inflammation (shortened colon, enlarged spleen) and increased adiposity (Fig. 3i, j and Extended Data Fig. 7u–a′). For P80, as little as 0.1% resulted in evidence of low-grade inflammation and increased adiposity, while 0.5% resulted in mild dysglycaemia (Fig. 3k, l and Extended Data Fig. 7b′–h′). Emulsifier-induced metabolic syndrome was also observed in older mice (Extended Data Fig. 5a–g) and persisted for at least 6 weeks after stopping emulsifier consumption (Extended Data Fig. 8a–k).

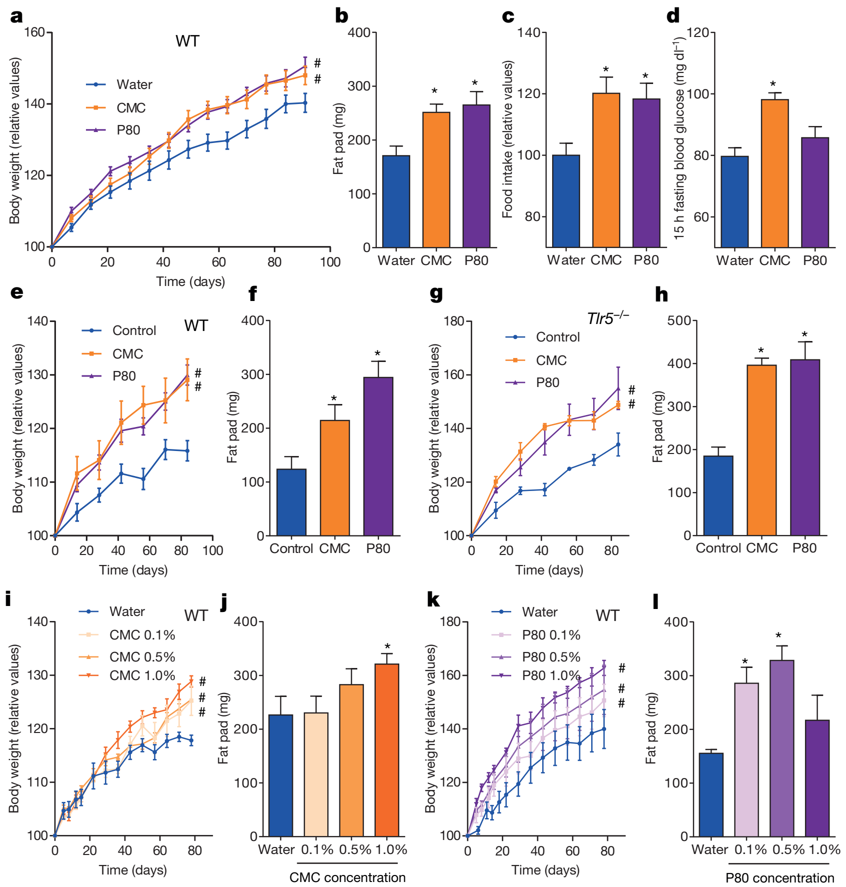


Figure 3 | Dietary emulsifiers promote metabolic syndrome. **a**–**d**, Wild-type (WT) mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **a**, Body weight over time; **b**, fat-pad mass; **c**, food intake; and **d**, 15 h fasting blood glucose concentration. **e**–**h**, Wild type (**e**–**f**) and *Tlr5*$^{-/-}$ (**g**–**h**) mice were given mouse chow containing CMC or P80 (1.0%) for 12 weeks. **e**, **g**, Body weights over time; **f**, **h**, fat-pad mass. **i**–**l**, Wild-type mice were exposed to drinking water containing 0.1–1.0% CMC (**i**, **j**) or P80 (**k**, **l**) for 12 weeks. **i**, **k**, Body weight over time; **j**, **l**, fat-pad mass. Data are the means ±s.e.m. ($n = 20$ for **a**–**d**, $n = 5$ for **e**–**l**). Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test (*$P < 0.05$) or two-way group ANOVA corrected for multiple comparisons with a Bonferroni test (#$P < 0.05$) compared to control group.

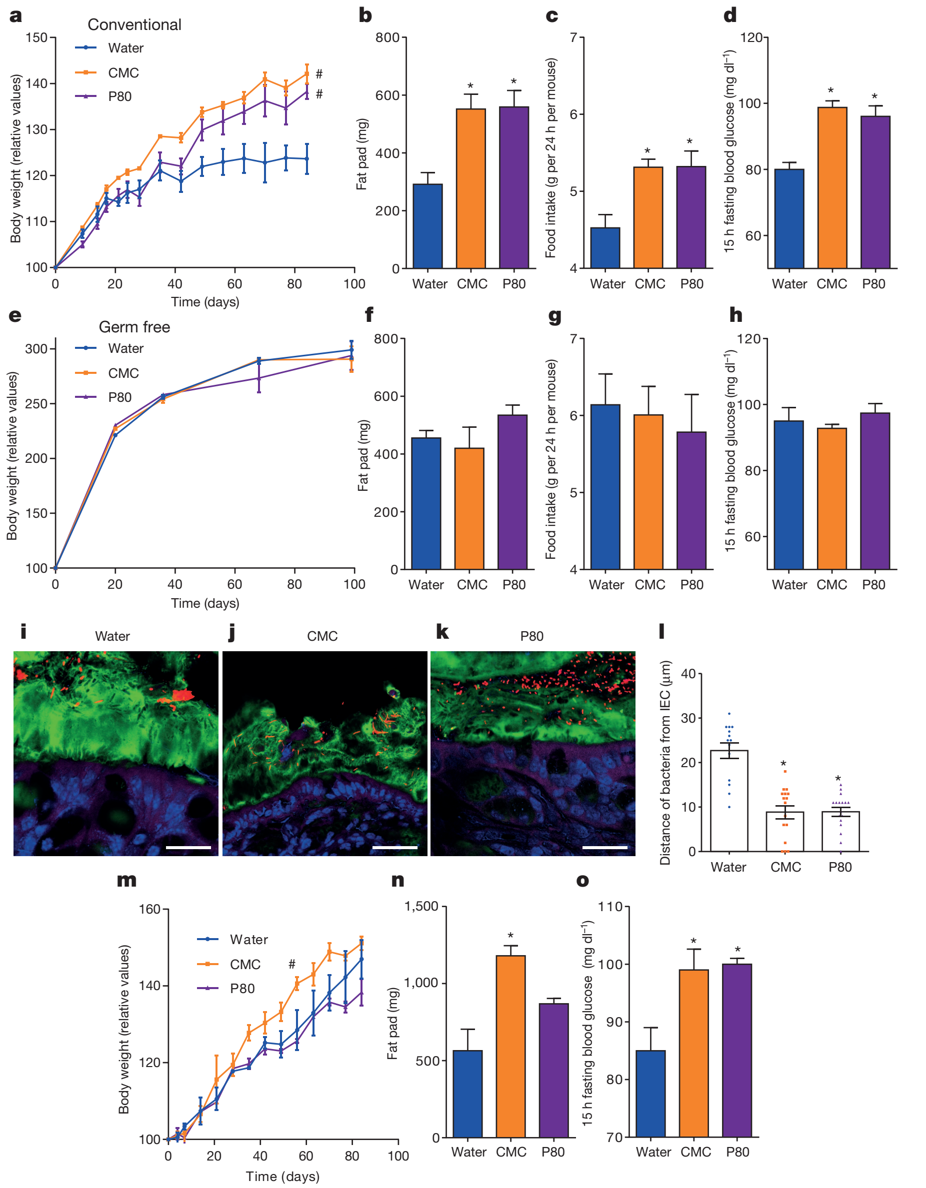


Figure 4 | Altered microbiota is necessary and sufficient for emulsifier-induced metabolic syndrome. **a**–**h**, Conventionally housed (**a**–**d**) and germ-free (**e**–**h**) Swiss Webster mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **a**, **e**, Body weight over time; **b**, **f**, fat-pad mass; **c**, **g**, food intake; and **d**, **h**, 15 h fasting blood glucose concentration. **i**–**o**, Germ-free Swiss Webster mice were conventionalized via microbiota transplant from mice that received standard drinking water or drinking water containing CMC or P80 (1.0%). **i**–**k**, Confocal microscopy analysis of microbiota localization: MUC2, green; actin, purple; bacteria, red; and DNA, blue. Scale bar, 20 μm. Pictures are representative of 10 biological replicates. **l**, Distances of closest bacteria to intestinal epithelial cells (IEC) per condition over five high-powered fields per mouse. **m**, Body weight over time; **n**, fat-pad mass; and **o**, 15 h fasting blood glucose concentration. Data are the means ± s.e.m. ($n = 5$ for **a**–**l**, $n = 3$ for **m**–**o**). Points are from individual mice. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test (*$P < 0.05$) or two-way group ANOVA corrected for multiple comparisons with a Bonferroni test (#$P < 0.05$) compared to control group.

Promotion of metabolic syndrome was not seen upon exposure to sodium sulfite, which is a common food additive but not an emulsifier (Extended Data Fig. 8l–s). *Tlr5*$^{-/-}$ mice are prone to developing metabolic syndrome[27], which results from poor microbiota management[18]. Emulsifiers markedly promoted multiple parameters of metabolic syndrome in *Tlr5*$^{-/-}$ mice (Extended Data Figs 7j–k and 8t–w), including hyperphagia, increased adiposity and glucose dysregulation, and was observed upon supplementation of chow or drinking water with as little as 0.1% P80 (Fig. 3g, h and Extended Data Fig. 8x–o′). A trend towards the development of metabolic syndrome in emulsifier-treated non-colitic *Il10*$^{-/-}$ mice was also observed, which was particularly evident upon exclusion of colitic mice, which exhibited weight loss (Extended Data Fig. 9a–f). Emulsifier-induced metabolic syndrome was observed in mice from multiple vivaria and strains, including Swiss Webster mice, which, in contrast to C57BL/6 mice, are considered obesity-resistant[28] (Fig. 4a–d and Extended Data Fig. 9g–k, x–a′).

Emulsifier-induced low grade inflammation and metabolic syndrome required the presence of a microbiota in that administration of emulsifiers to germ-free mice resulted in neither low-grade inflammation, as assessed by faecal LCN2, colon length and splenomegaly, nor all measured parameters of metabolic syndrome including body mass, fat mass, food intake and fasting glucose levels (Fig. 4e–h and Extended Data Fig. 9g–p). This result could reflect direct action of CMC and P80 on gut bacteria composition or metabolism. In accord with both of these possibilities, emulsifiers altered faecal levels of short-chain fatty acids (Extended Data Fig. 9q–w), including decreased levels of butyrate, which is thought to play a key role in dampening inflammation[29]. Emulsifiers also altered bile acid levels (Extended Data Fig. 9b′–t′), which influence microbiota composition[24]. However, in germ-free mice, emulsifiers did not alter bile acids, reduce mucus thickness nor alter penetrance of 0.5 μm beads into the mucus (Extended Data Figs 9b′–t′ and 10a–l), suggesting that these changes are not purely a direct effect of emulsifiers on mucus structures. Collectively, these data suggest that alterations in mucus in emulsifier-treated mice might result, at least in part, from altered gut microbiota composition but do not exclude direct effects on the host.

Transfer of microbiota from emulsifier-treated mice to germ-free mice (not exposed to emulsifiers) transferred microbiota encroachment, low-grade inflammation, increased adiposity and dysglycaemia (Fig. 4i–o and Extended Data Fig. 10s–y). Microbiota transplant from CMC-treated mice resulted in a longer-lasting increase in adiposity, paralleling that adiposity was less rapidly reversible following stoppage of CMC exposure (Extended Data Fig. 8a–k). Such transplanted phenotypes correlated with the acquisition of elevated levels of faecal LPS and flagellin

(Extended Data Fig. 10m–r) and altered microbiota composition (Extended Data Fig. 10z–e′). Thus, emulsifier-induced changes in the microbiota have a role in driving the inflammation and metabolic changes promoted by these food additives.

The last half-century has witnessed a steady increase in the consumption of food additives, many of which have not been carefully tested as they were given GRAS status at the time that government entities charged with regulating food safety were created and/or expanded. Moreover, the testing of food additives that has been performed has generally used animal models designed to detect acute toxicity and/or promotion of cancer. Our data suggest that such testing may be inadequate—a notion supported by the recent observation that artificial sweeteners induce dysglycaemia in humans[30]. More specifically, our data suggest that one ubiquitous class of food additives, namely emulsifiers, can disturb the host–microbiota relationship resulting in a microbiota with enhanced mucolytic and pro-inflammatory activity that promotes intestinal inflammation. Such chronic gut inflammation can manifest as colitis or metabolic syndrome. While additional studies will be needed to determine if CMC, P80 and/or other emulsifiers impact human health, our observations in mice suggest the possibility that dietary emulsifiers may have contributed to the post-mid-twentieth-century increase in incidence of inflammatory bowel disease, metabolic syndrome, and perhaps other chronic inflammatory diseases. Notably, this hypothesis does not dispute the commonly held assumption that excess caloric consumption is a predominant factor driving the metabolic syndrome epidemic. Rather, it suggests such hyperphagia may be driven, in part, by food additives and other factors that might alter gut microbiota and promote low-grade intestinal inflammation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 December 2013; accepted 14 January 2015.
Published online 25 February 2015.

1. Johansson, M. E. *et al.* The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc. Natl Acad. Sci. USA* **105,** 15064–15069 (2008).
2. Roberts, C. L. *et al.* Translocation of Crohn's disease *Escherichia coli* across M-cells: contrasting effects of soluble plant fibres and emulsifiers. *Gut* **59,** 1331–1339 (2010).
3. Swidsinski, A., Loening-Baucke, V. & Herber, A. Mucosal flora in Crohn's disease and ulcerative colitis - an overview. *J. Physiol. Pharmacol.* **60** (Suppl 6), 61–71 (2009).
4. Food Safety Commission [of Japan]. Evaluation Report of Food Additives (Polysorbates 20, 60, 65 and 80) (2007) [transl.] https://www.fsc.go.jp/english/evaluationreports/foodadditive/polysorbate_report.pdf.
5. National Toxicology Program. Toxicology and Carcinogenesis Studies of Polysorbate 80 (CAS No. 9005–65–6) in F344/N Rats and B6C3F1 Mice (Feed Studies). *Natl. Toxicol. Program Tech. Rep. Ser.* **415,** 1–225 (1992).
6. Swidsinski, A. *et al.* Bacterial overgrowth and inflammation of small intestine after carboxymethylcellulose ingestion in genetically susceptible mice. *Inflamm. Bowel Dis.* **15,** 359–364 (2009).
7. Kühn, R., Lohler, J., Rennick, D., Rajewsky, K. & Muller, W. Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* **75,** 263–274 (1993).
8. Vijay-Kumar, M. *et al.* Deletion of TLR5 results in spontaneous colitis in mice. *J. Clin. Invest.* **117,** 3909–3921 (2007).
9. Ubeda, C. *et al.* Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *J. Exp. Med.* **209,** 1445–1456 (2012).
10. Chassaing, B., Ley, R. E. & Gewirtz, A. T. Intestinal epithelial cell Toll-like receptor 5 regulates the intestinal microbiota to prevent low-grade inflammation and metabolic syndrome in mice. *Gastroenterology* **147,** 1363–1377 (2014).
11. Johansson, M. E. & Hansson, G. C. Preservation of mucus in histological sections, immunostaining of mucins in fixed tissue, and localization of bacteria with FISH. *Methods Mol. Biol.* **842,** 229–235 (2012).
12. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71,** 8228–8235 (2005).
13. Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science* **307,** 1915–1920 (2005).
14. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).
15. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
16. Png, C. W. *et al.* Mucolytic bacteria with increased prevalence in IBD mucosa augment *in vitro* utilization of mucin by other bacteria. *Am. J. Gastroenterol.* **105,** 2420–2428 (2010).
17. Arthur, J. C. *et al.* Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338,** 120–123 (2012).
18. Carvalho, F. A. *et al.* Transient inability to manage proteobacteria promotes chronic gut inflammation in TLR5-deficient mice. *Cell Host Microbe* **12,** 139–152 (2012).
19. Mukhopadhya, I., Hansen, R., El-Omar, E. M. & Hold, G. L. IBD-what role do Proteobacteria play? *Nature Rev. Gastroenterol. Hepatol.* **9,** 219–230 (2012).
20. Cox, L. M. *et al.* Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell* **158,** 705–721 (2014).
21. Chassaing, B., Koren, O., Carvalho, F. A., Ley, R. E. & Gewirtz, A. T. AIEC pathobiont instigates chronic colitis in susceptible hosts by altering microbiota composition. *Gut* **63,** 1069–1080 (2014).
22. Ziegler, T. R. *et al.* Detectable serum flagellin and lipopolysaccharide and upregulated anti-flagellin and lipopolysaccharide immunoglobulins in human short bowel syndrome. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **294,** R402–R410 (2008).
23. Chassaing, B. *et al.* Fecal lipocalin 2, a sensitive and broadly dynamic non-invasive biomarker for intestinal inflammation. *PLoS ONE* **7,** e44328 (2012).
24. Devkota, S. *et al.* Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in $Il10^{-/-}$ mice. *Nature* **487,** 104–108 (2012).
25. Kullberg, M. C. *et al. Helicobacter hepaticus* triggers colitis in specific-pathogen-free interleukin-10 (IL-10)-deficient mice through an IL-12- and gamma interferon-dependent mechanism. *Infect. Immun.* **66,** 5157–5166 (1998).
26. Gregor, M. F. & Hotamisligil, G. S. Inflammatory mechanisms in obesity. *Annu. Rev. Immunol.* **29,** 415–445 (2011).
27. Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328,** 228–231 (2010).
28. Wong, T. *et al.* Divergent metabolic adaptations to intestinal parasitic nematode infection in mice susceptible or resistant to obesity. *Gastroenterology* **133,** 1979–1988 (2007).
29. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504,** 446–450 (2013).
30. Suez, J. *et al.* Artificial sweeteners induce glucose intolerance by altering the gut microbiota. *Nature* **514,** 181–186 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by NIH grant DK099071 and DK083890. B.C. is a recipient of the Research Fellowship award from the Crohn's and Colitis Foundation of America (CCFA). We thank B. Zhang, L. Etienne-Mesmin, H. Q. Tran and E. Viennois for technical assistance.

Author Contributions B.C. and A.T.G. conceived the project, designed the experiments, interpreted the results, and wrote the manuscript. B.C. performed all experiments and analysis with advice and guidance from O.K., J.K.G., and A.C.P. S.S. and R.E.L. guided experimental design and data interpretation.

Author Information Sequencing data are deposited in the European Nucleotide Archive under accession number PRJEB8035. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.T.G. (agewirtz@gsu.edu).

METHODS

Mice. Wild-type C57BL/6 and *Il10*$^{-/-}$ mice were purchased from Jackson Laboratories. Swiss Webster mice were purchased from Taconic, Inc. *Tlr5*$^{-/-}$ mice were originally generated by S. Akira (Osaka University, Japan)[31] and backcrossed/maintained as previously described[27]. All mice were then bred and housed at Georgia State University, Atlanta, Georgia, USA under institutionally-approved protocols (Institutional Animal Care and Use Committee no. A14033). For wild-type mice, experiments were also performed on mice purchased directly from the aforementioned providers and used within 3 days of receipt. Mice were fed Purina rodent chow no. 5001, which is commonly used in many vivaria. Mice were weaned at 3 weeks of age and put into new cages (randomized to littermates) such that each experimental group contained mice from at least two different litters, and that each litter were used for at least three experimental groups. Experiments used either male or female mice (comparisons within a gender) except for Figs 1m–o and 2 which used both male and female mice. For littermate control experiments (Extended Data Fig 3), all the mice from 10 (Extended Data Fig. 3a–g) and 8 (Extended Data Fig. 3h–n) different litters (from distinct mothers) were used and placed into cages such that each litter was split equally amongst groups that were to receive water, CMC or P80 (three (Extended Data Fig. 3a–g) or two (Extended Data Fig. 3h–n) cages per sex per condition). For experiments using older mice (Extended Data Fig. 5a–p), mice were weaned at 3 weeks of age, put into new cages, and subjected to emulsifier treatment starting 3 months later.

Materials. Sodium carboxymethylcellulose (CMC, average M_w ~250,000), polysorbate-80 (P80) and sodium sulfite were purchased from Sigma (Sigma, St. Louis, Missouri).

Emulsifier agent treatment. Mice were exposed to CMC, P80 or sodium sulfite diluted in drinking water (1.0% or indicated concentration) (not blinded). The same water (reverse-osmosis treated Atlanta city water) was used for the water-treated (control) group. These solutions were changed every week. Emulsifier solutions were autoclaved for experiments presented in Fig. 4a–h. When required, emulsifier solutions were replaced by water after 6 weeks of treatment for an additional 6 weeks. Body weights were measured every week and are expressed as a percentage compared to the initial body weight (day 0) defined as 100%. Mice developing colitis (LCN2 level ≥ 500 ng per g of faeces) were excluded for metabolic syndrome parameters analysis. Fresh faeces were collected every week for downstream analysis. After 3 months of emulsifier treatment, mice were fasted for 5 h, at which time blood was collected by retrobulbar intraorbital capillary plexus. Haemolysis-free serum was generated by centrifugation of blood using serum separator tubes (Becton Dickinson, Franklin Lakes, New Jersey). Mice were then euthanized and colon length, colon weight, spleen weight and adipose weight were measured. Organs were collected for downstream analysis. Incorporation of CMC and P80 into the diet was performed by Research Diets Inc. (New Brunswick, New Jersey) using Purina Rodent Chow diet no. 5001. These diets are now referenced by Research Diets as C13050701 and C13050702, respectively.

Food intake measurement. Groups of mice were placed in a clean cage with a known amount of food. Twenty-four hours later, the amount of remaining food was measured with the difference viewed as food intake per 24 h. Error bars represent s.e.m. of three measurements made one week apart.

Overnight fasting blood glucose measurement. Mice were placed in a clean cage and fasted for 15 h. Blood glucose concentration was then determined using a Nova Max Plus Glucose Meter and expressed in mg dl^{-1}.

Glucose and insulin tolerance test. Wild-type and *Tlr5*$^{-/-}$ mice were treated with CMC or P80 diluted to 1% drinking water for 8 weeks. Following a 5-h fasting, baseline blood glucose levels were measured using a Nova Max Plus Glucose Meter and expressed in mg dl^{-1}. Mice were then injected intraperitoneally with 2 g glucose per kg body weight in sterile PBS or with 0.5 U insulin per kg body weight (Sigma, St. Louis, Missouri), and blood glucose levels were measured 30, 60, 90 and 120 min after injection, as previously described[27].

***In vivo* epithelial barrier permeability.** *In vivo* assay of intestinal barrier function was performed using an FITC-labelled dextran method, as previously described[32]. Mice were deprived of food and water for 4 h, and were then gavaged with 15 mg of permeability tracer FITC-labelled dextran 4 kDa (FD4, Sigma, St. Louis, Missouri). Blood was collected retro-orbitally after 3 h, and fluorescence intensity was measured in the serum (excitation, 490 nm; emission, 520 nm; BIOTEK Fluorescence Spectrophotometer). FITC–dextran concentrations were determined using a standard curve generated by serial dilution of FITC–dextran in mice serum.

Germ-free experiments. Germ-free Swiss Webster mice were kept under germ-free conditions in a Park Bioservices isolator in our germ-free facility. CMC and P80 were diluted to 1% in water and then autoclaved for germ-free purpose. The same water was used for the water-treated (control) group. After 3 months of emulsifier agent treatment, mice were fasted for 5 h and then removed from the isolator to be euthanized. Samples were collected as previously described. For the analysis of mucus layer integrity, germ-free C57BL/6 mice were kept under germ-free conditions and treated with CMC or P80 as described above. After 2 months of emulsifier agent treatment, mice were removed from the isolator, and inoculated with 0.5 μm green fluorescent beads (Polysciences, Warrington, PA). Seven hours post-inoculation, mice were euthanized and colonic tissue was collected and placed in methanol-Carnoy's fixative solution (60% methanol, 30% chloroform, 10% glacial acetic acid), followed by immunostaining of mucins, as described below.

Microbiota transplantation. Caecal contents from detergent treated mice were suspended in 30% glycerol diluted in PBS (1.0 ml) and stocked at −80 °C until analysis. Germ-free Swiss Webster mice (4 weeks old) were removed from the isolator and were orally administered 200 μl of faecal suspension made using glycerol stocks. Transplanted mice were then monitored as previously described.

Colonic myeloperoxidase assay. Neutrophil influx in tissue was analysed by assaying the enzymatic activity of myeloperoxidase, a marker for neutrophils. In brief, tissue (50 mg ml^{-1}) was thoroughly washed in PBS and homogenized in 0.5% hexadecyltrimethylammonium bromide (Sigma, St. Louis, Missouri) in 50 mM PBS (pH 6.0), freeze–thawed three times, sonicated and centrifuged. Myeloperoxidase was assayed in the clear supernatant by adding 1 mg ml^{-1} of dianisidine dihydrochloride (Sigma, St. Louis, Missouri) and 5×10^{-4}% H_2O_2 and the change in optical density measured at 450 nm. Human neutrophil myeloperoxidase (Sigma, St. Louis, Missouri) was used as a standard. One unit of myeloperoxidase activity was defined as the amount that degraded 1.0 μmol of peroxide per min at 25 °C[33].

Haematoxylin and eosin staining and histopathologic analysis. Following euthanasia, mouse colons and small intestines were fixed in 10% buffered formalin for 24 h at room temperature and then embedded in paraffin. Tissues were sectioned at 5 μm thickness and stained with haematoxylin and eosin (H&E) using standard protocols. H&E-stained slides were scored as follows. Each colon was assigned four scores based on the degree of epithelial damage and inflammatory infiltrate in the mucosa, submucosa and muscularis/serosa, as previously described[34]. A slight modification was made to this scoring system[23]; each of the four scores was multiplied by 1 if the change was focal, 2 if it was patchy and 3 if it was diffuse. The four individual scores per colon were added, resulting in a total scoring range of 0–36 per mouse.

Periodic acid–Schiff staining. Following euthanasia, mouse colons were fixed in 10% buffered formalin for 24 h at room temperature and then embedded in paraffin. Tissues were sectioned at 5 μm thickness, deparaffinized and oxidized for 5 min in 0.5% periodic acid solution (Sigma, St. Louis, Missouri). After rinsing, tissues were placed in Schiff reagent (Sigma, St. Louis, Missouri) for 15 min, washed in warm water, and counterstained using haematoxylin (Sigma, St. Louis, Missouri) for one minute.

Immunostaining of mucins and localization of bacteria by fluorescent *in situ* hybridization. Mucus immunostaining was paired with fluorescent *in situ* hybridization (FISH), as previously described[11], in order to analyse bacteria localization at the surface of the intestinal mucosa. In brief, colonic tissues (proximal colon, second cm from the caecum) containing faecal material were placed in methanol-Carnoy's fixative solution (60% methanol, 30% chloroform, 10% glacial acetic acid) for a minimum of 3 h at room temperature. Tissue were then washed in methanol 2 × 30 min, ethanol 2 × 15 min, ethanol/xylene (1:1) 15 min and xylene 2 × 15 min, followed by embedding in paraffin with a vertical orientation. Five-μm sections were cut and dewaxed by preheating at 60 °C for 10 min, followed by bathing in xylene at 60 °C for 10 min, xylene at room temperature for 10 min and 99.5% ethanol for 10 min. The hybridization step was performed at 50 °C overnight with an EUB338 probe (5′-GCTGCCTCCCGTAGGAGT-3′, with a 5′ Alexa 647 label) diluted to a final concentration of 10 μg ml^{-1} in hybridization buffer (20 mM Tris-HCl, pH 7.4, 0.9 M NaCl, 0.1% SDS, 20% formamide). After washing for 10 min in wash buffer (20 mM Tris-HCl, pH 7.4, 0.9 M NaCl) and 3 × 10 min in PBS, a PAP pen (Sigma, St. Louis, Missouri) was used to mark around the section and block solution (5% FBS in PBS) was added for 30 min at 4 °C. Mucin 2 primary antibody (rabbit H-300, Santa Cruz Biotechnology) was diluted to 1:1500 in block solution and applied overnight at 4 °C. After washing 3 × 10 min in PBS, block solution containing anti-rabbit Alexa 488 secondary antibody diluted to 1:1500, Phalloidin–Tetramethylrhodamine B isothiocyanate (Sigma, St. Louis, Missouri) at 1 μg ml^{-1} and Hoechst 33258 (Sigma, St. Louis, Missouri) at 10 μg ml^{-1} was applied to the section for 2 h. After washing 3 × 10 min in PBS slides were mounted using Prolong anti-fade mounting media (Life Technologies). Observations were performed with a Zeiss LSM 700 confocal microscope with software Zen 2011 version 7.1. This software was used to determine the distance between bacteria and the epithelial cell monolayer, as well as the mucus thickness.

Quantification of faecal LCN2 by ELISA and determination of colitis incidence. For quantification of faecal LCN2 by ELISA, frozen faecal samples were reconstituted in PBS containing 0.1% Tween 20 to a final concentration of 100 mg ml^{-1} and vortexed for 20 min to produce a homogenous faecal suspension[23]. These samples were then centrifuged for 10 min at 14,000*g* and 4 °C. Clear supernatants were collected and stored at −20 °C until analysis. LCN2 levels were estimated in the

supernatants using Duoset murine LCN2 ELISA kit (R&D Systems, Minneapolis, Minnesota) using the colourimetric peroxidase substrate tetramethylbenzidine, and optical density was read at 450 nm (Versamax microplate reader). For determination of colitis incidence, a faecal LCN2 level ≥ 500 ng per g of faeces was used to determine colitic mice.

Faecal flagellin and LPS load quantification. We quantified flagellin and LPS as previously described[21] using human embryonic kidney (HEK)-Blue-mTLR5 and HEK-Blue-mTLR4 cells, respectively (Invivogen, San Diego, California). We resuspended faecal material in PBS to a final concentration of 100 mg ml^{-1} and homogenized for 10 s using a Mini-Beadbeater-24 without the addition of beads to avoid bacteria disruption. We then centrifuged the samples at 8000*g* for 2 min and serially diluted the resulting supernatant and applied to mammalian cells. Purified *Escherichia coli* flagellin and LPS (Sigma, St Louis, Missouri) were used for standard curve determination using HEK-Blue-mTLR5 and HEK-Blue-mTLR4 cells, respectively. After 24 h of stimulation, we applied cell culture supernatant to QUANTI-Blue medium (Invivogen, San Diego, California) and measured alkaline phosphatase activity at 620 nm after 30 min.

Serum flagellin- and LPS-specific immunoglobulins. Flagellin- and LPS-specific IgG levels were quantified by ELISA, as previously described[22,35]. Microtitre plates were coated overnight with purified *E. coli* flagellin (100 ng per well) or LPS (1 μg per well). Serum samples diluted 1:500 were then applied. After incubation and washing, wells were incubated with anti-mouse IgG. Quantification was performed using the colourimetric peroxidase substrate tetramethylbenzidine, as described above. Data are reported as optical density corrected by subtracting background (determined by readings in samples lacking serum).

Gene expression analysis by quantitative reverse-transcription PCR (qRT–PCR). Total RNAs were isolated from colonic tissues using TRIzol (Invitrogen, Carlsbad, California) according to the manufacturer's instructions. Quantitative RT–PCR was performed using the iScriptTM One-Step RT–PCR Kit with SYBR Green (Bio-Rad, Hercules, Californai) in a CFX96 apparatus (Bio-Rad, Hercules, California) with specific mouse oligonucleotides. The oligonucleotides used were: *36B4* (sense) 5′-TCCAGGCTTTGGGCATCA-3′ and (antisense) 5′-CTTTATTCAGCTGCACATCACTCAGA-3′, *Tff3* (sense) 5′-CCTGGTTGCTGGGTCCTCTGG-3′ and (antisense) 5′-GTCTCCTGCAGAGGTTTGAAGC-3′, *Klf4F* (sense) 5′-TGTGACTATGCAGGCTGTGG-3′ and (antisense) 5′-AGTGCCTGGTCAGTTCATCG-3′. Results were normalized to the housekeeping *36B4* gene.

Bacterial quantification by qPCR. For quantification of total faecal bacterial load, total bacterial DNA was isolated from weighted faeces using QIAamp DNA Stool Mini Kit (Qiagen). DNA was then subjected to quantitative PCR using QuantiFast SYBR Green PCR kit (Biorad) with universal 16S rRNA primers 8F: 5′-AGAGTTTGATCCTGGCTCAG-3′ and 338R: 5′-CTGCTGCCTCCCGTAGGAGT-3′ to measure total bacteria[36]. Results are expressed as bacteria number per mg of stool, using a standard curve. For quantification of mucosa-associated bacteria, total DNA was isolated from PBS-washed and weighted colonic tissue using DNeasy Blood & Tissue Kit (Qiagen). DNA was then subjected to quantitative PCR as described above, and results are expressed as bacteria number per mg of colonic tissue, using a standard curve.

Faecal microbiota analysis by 16S rRNA gene sequencing using Illumina technology. 16S rRNA gene amplification and sequencing were done using the Illumina MiSeq technology following the protocol of Earth Microbiome Project with their modifications to the MO BIO PowerSoil DNA Isolation Kit procedure for extracting DNA (http://www.earthmicrobiome.org/emp-standard-protocols)[37,38]. Bulk DNA were extracted from frozen extruded faeces using a PowerSoil-htp kit from MO BIO Laboratories (Carlsbad, California) with mechanical disruption (bead beating). The 16S rRNA genes, region V4, were PCR amplified from each sample using a composite forward primer and a reverse primer containing a unique 12-base barcode, designed using the Golay error-correcting scheme, which was used to tag PCR products from respective samples[38]. We used the forward primer 515F 5′-*AATGATACGGCGACCACCGAGATCTACAC***TATGGTAATT*****GT***GTGCCAGCMGCCGCGGTAA-3′; the italicized sequence is the 5′ Illumina adaptor B, the bold sequence is the primer pad, the italicized and bold sequence is the primer linker and the underlined sequence is the conserved bacterial primer 515F. The reverse primer 806R used was 5′-*CAAGCAGAAGACGGCATACGAGAT* XXXXXXXXXXXX **AGTCAGTCAG** ***CC*** GGACTACHVGGGTWTCTAAT-3′; the italicized sequence is the 3′ reverse complement sequence of Illumina adaptor, the 12 X sequence is the Golay barcode, the bold sequence is the primer pad, the italicized and bold sequence is the primer linker and the underlined sequence is the conserved bacterial primer 806R. PCR reactions consisted of Hot Master PCR mix (Five Prime), 0.2 μM of each primer, 10–100 ng template, and reaction conditions were 3 min at 95 °C, followed by 30 cycles of 45 s at 95 °C, 60 s at 50 °C and 90 s at 72 °C on a Biorad thermocycler. Four independent PCRs were performed for each sample, combined, purified with Ampure magnetic purification beads (Agencourt), and products were visualized by gel electrophoresis. Products were then quantified (BIOTEK Fluorescence Spectrophotometer) using a Quant-iT PicoGreen dsDNA assay. A master DNA pool was generated from the purified products in equimolar ratios. The pooled products were quantified using a Quant-iT PicoGreen dsDNA assay and then sequenced using an Illumina MiSeq sequencer (paired-end reads, 2 × 250 base pairs) at Cornell University, Ithaca.

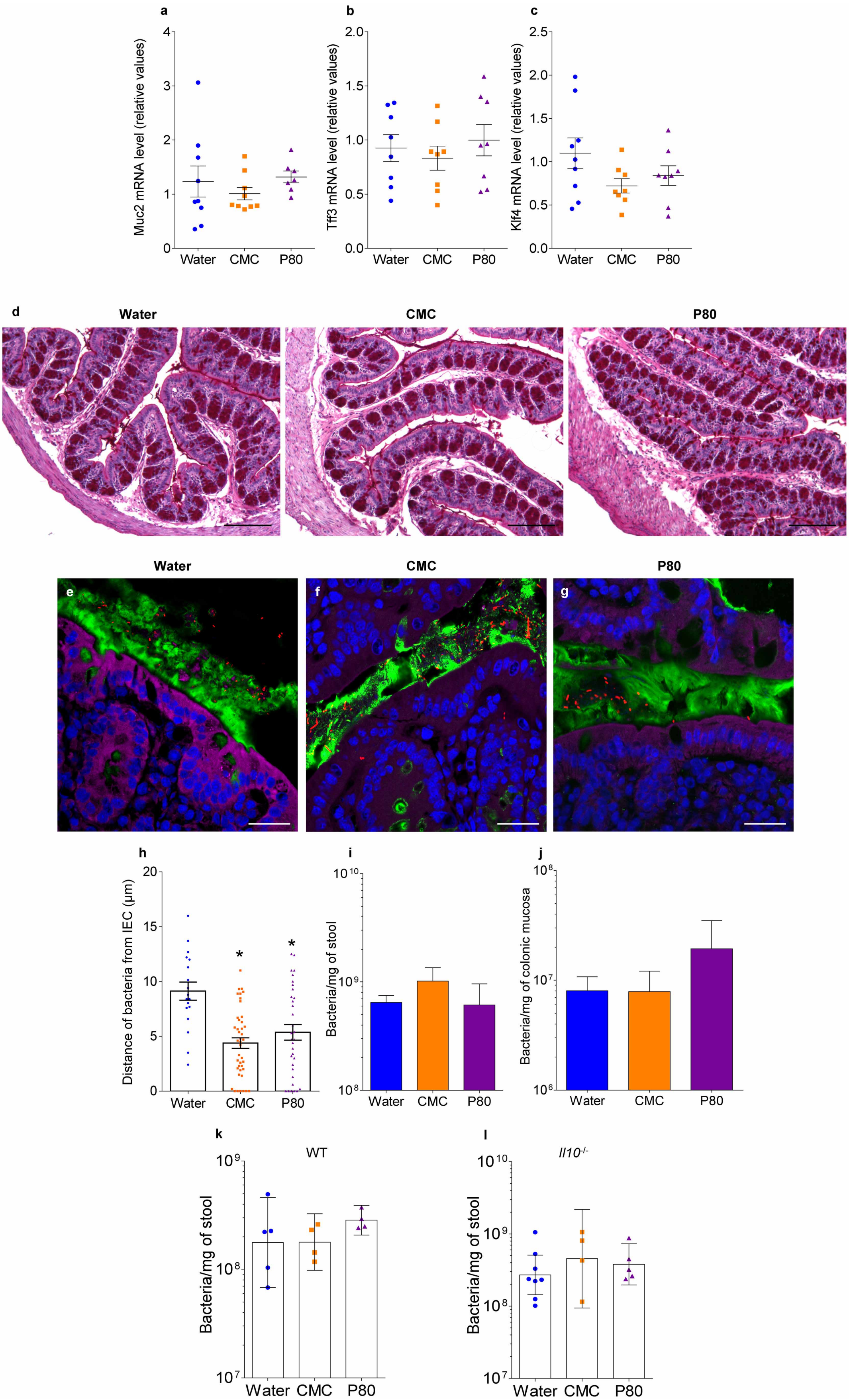
16S rRNA gene sequence analysis. The sequences were demultiplexed, quality filtered using the Quantitative Insights Into Microbial Ecology (QIIME, version 1.8.0) software package[39], and forward and reverse Illumina reads were joined using the fastq-join method (http://code.google.com/p/ea-utils)[40]. We used QIIME default parameters for quality filtering (reads truncated at first low-quality base and excluded if: (1) there were more than three consecutive low quality base calls; (2) less than 75% of read length was consecutive high quality base calls; (3) at least one uncalled base was present; (4) more than 1.5 errors were present in the barcode; (5) any Phred qualities were below 20; or (6) the length was less than 75 bases). Sequences were assigned to OTUs using the UCLUST algorithm[41] with a 97% threshold of pairwise identity (without the creation of new clusters with sequences that do not match the reference sequences), and classified taxonomically using the Greengenes reference database[42]. A single representative sequence for each OTU was aligned and a phylogenetic tree was built using FastTree[43]. The phylogenetic tree was used for computing the unweighted UniFrac distances between samples[12,44], rarefaction were performed (3,500–20,000 sequences per sample) and used to compare abundances of OTUs across samples. Principal coordinates analysis (PCoA) plots were used to assess the variation between experimental group (beta diversity) and jackknifed beta diversity was used to estimate the uncertainty in PCoA plots. In addition to using PCoA, samples were clustered using UPGMA (unweighted pair group method with arithmetic mean). Alpha diversity curves were determined for all samples using the determination of the number of observed species, and the Shannon diversity index was used to characterize species diversity in a community. Sequencing data are deposited in the European Nucleotide Archive under accession number PRJEB8035.

Short-chain fatty acids and bile acids composition analysis. Faecal samples were used to analyse short-chain fatty acids and bile acid composition at the Metabolomics Core of the University of Michigan (supported by grant U24 DK097153 of NIH Common Funds Project to the University of Michigan).

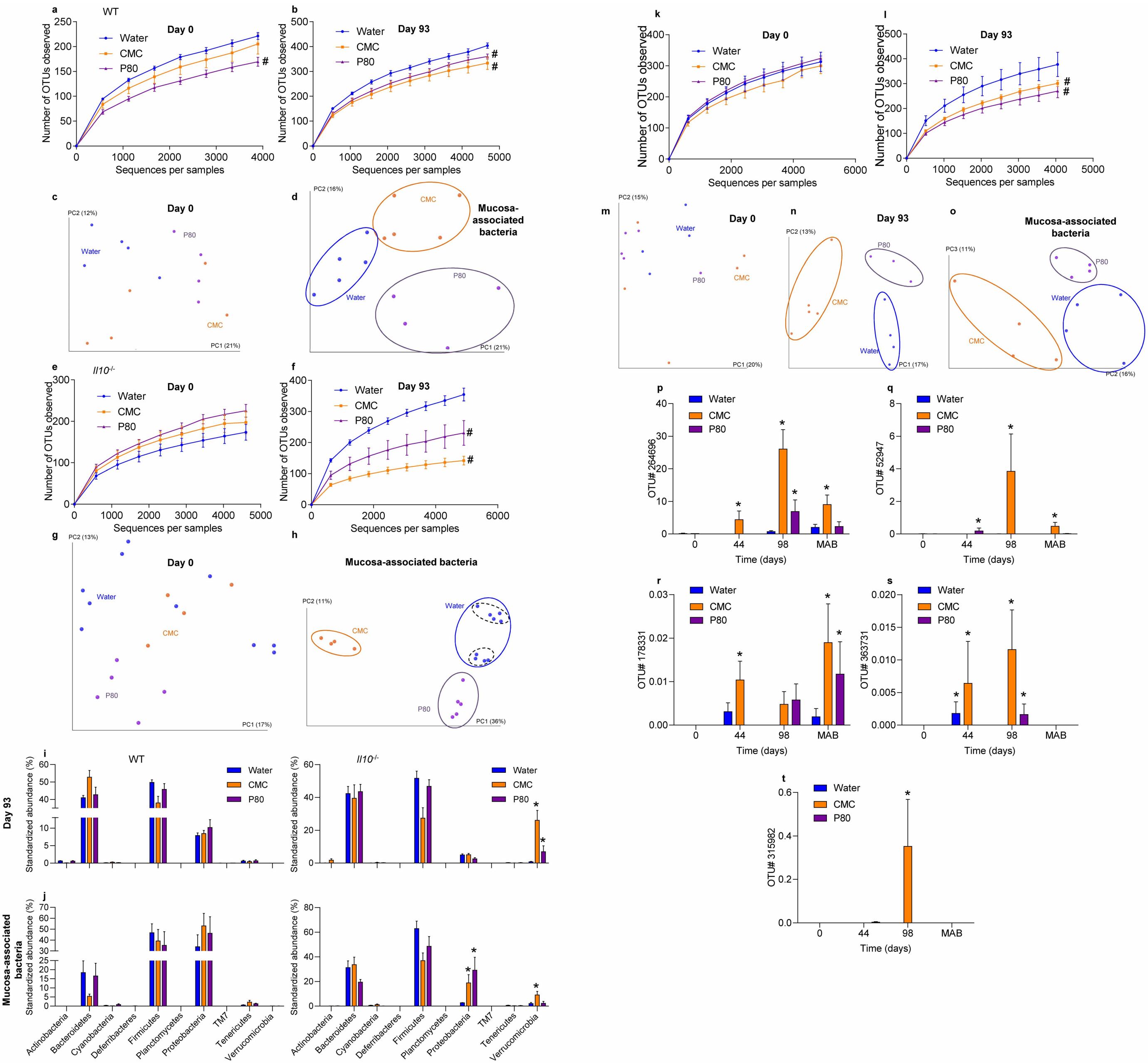
Statistical analysis. All replicates in this study were biological; that is, repeat experiments with additional mice. The D'Agostino–Pearson omnibus test was used to verify that all data were normally distributed. Significance was determined using *t*-tests, one-way ANOVA corrected for multiple comparisons with a Sidak test, two-way ANOVA corrected for multiple comparisons with a Bonferroni test, or two-way group ANOVA (GraphPad Prism software, version 6.01). Differences were noted as significant at $P \leq 0.05$. A 'nearest-shrunken centroid' classification approach was performed to detect the OTUs that were particularly representative of each experimental group[45]. The amount of shrinkage was set to minimize the misclassification error. When OTUs were used to classify samples, accuracy of the classification was perfect (overall error rate = 0). These analyses allowed the identification of the 15 OTUs whose abundances were the most significantly different between experimental groups, which were then used for heat map generation and sample clustering. These analyses were performed using the prediction analysis for microarrays (PAM) package within R software[45]. No statistical methods were used to predetermine sample size.

31. Uematsu, S. *et al.* Detection of pathogenic intestinal bacteria by Toll-like receptor 5 on intestinal CD11c^{+} lamina propria cells. *Nature Immunol.* **7,** 868–874 (2006).
32. Denizot, J. *et al.* Adherent-invasive *Escherichia coli* induce claudin-2 expression and barrier defect in CEABAC10 mice and Crohn's disease patients. *Inflamm. Bowel Dis.* **18,** 294–304 (2012).
33. Castaneda, F. E. *et al.* Targeted deletion of metalloproteinase 9 attenuates experimental colitis in mice: central role of epithelial-derived MMP. *Gastroenterology* **129,** 1991–2008 (2005).
34. Katakura, K. *et al.* Toll-like receptor 9-induced type I IFN protects mice from experimental colitis. *J. Clin. Invest.* **115,** 695–702 (2005).
35. Sitaraman, S. V. *et al.* Elevated flagellin-specific immunoglobulins in Crohn's disease. *Am. J. Physiol. Gastrointest. Liver Physiol.* **288,** G403–G406 (2005).
36. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).
37. Gilbert, J. A. *et al.* The Earth Microbiome Project: Meeting report of the ''1 EMP meeting on sample selection and acquisition'' at Argonne National Laboratory October 6 2010. *Stand. Genomic Sci.* **3,** 249–253 (2010).
38. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6,** 1621–1624 (2012).
39. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7,** 335–336 (2010).
40. Aronesty, E. Comparison of sequencing utility programs. *Open Bioinformatics J.* **7,** 1–8 (2013).
41. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).

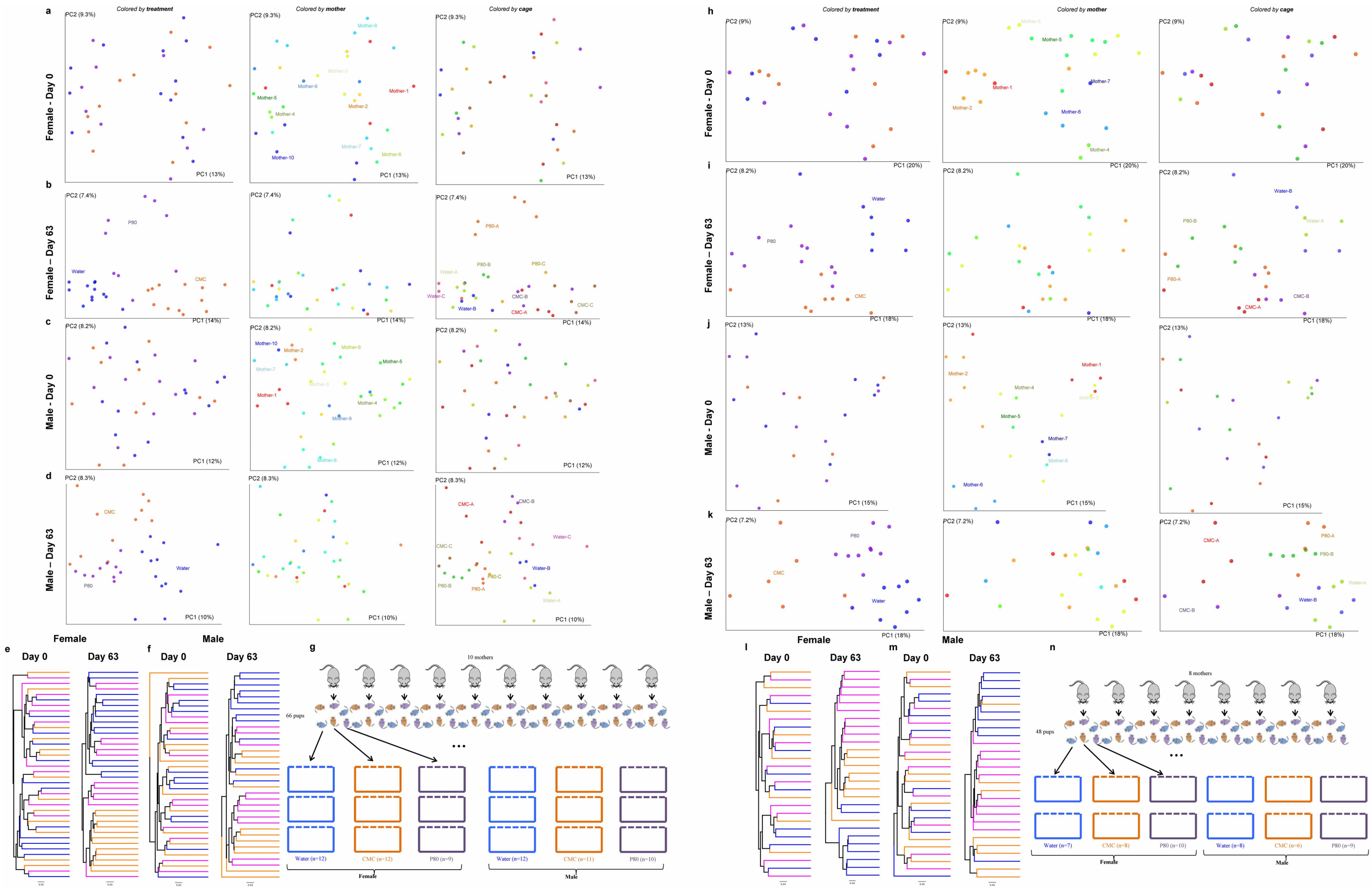
42. McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6,** 610–618 (2012).
43. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26,** 1641–1650 (2009).
44. Lozupone, C., Hamady, M. & Knight, R. UniFrac–an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7,** 371 (2006).
45. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl Acad. Sci. USA* **99,** 6567–6572 (2002).


a
Muc2 mRNA level (relative values)
Water
CMC
P80
b
Tff3 mRNA level (relative values)
c
Klf4 mRNA level (relative values)
d
Water
CMC
P80
e
f
g
h
Distance of bacteria from IEC (µm)
*
*
i
Bacteria/mg of stool
j
Bacteria/mg of colonic mucosa
k
WT
l
Il10-/-

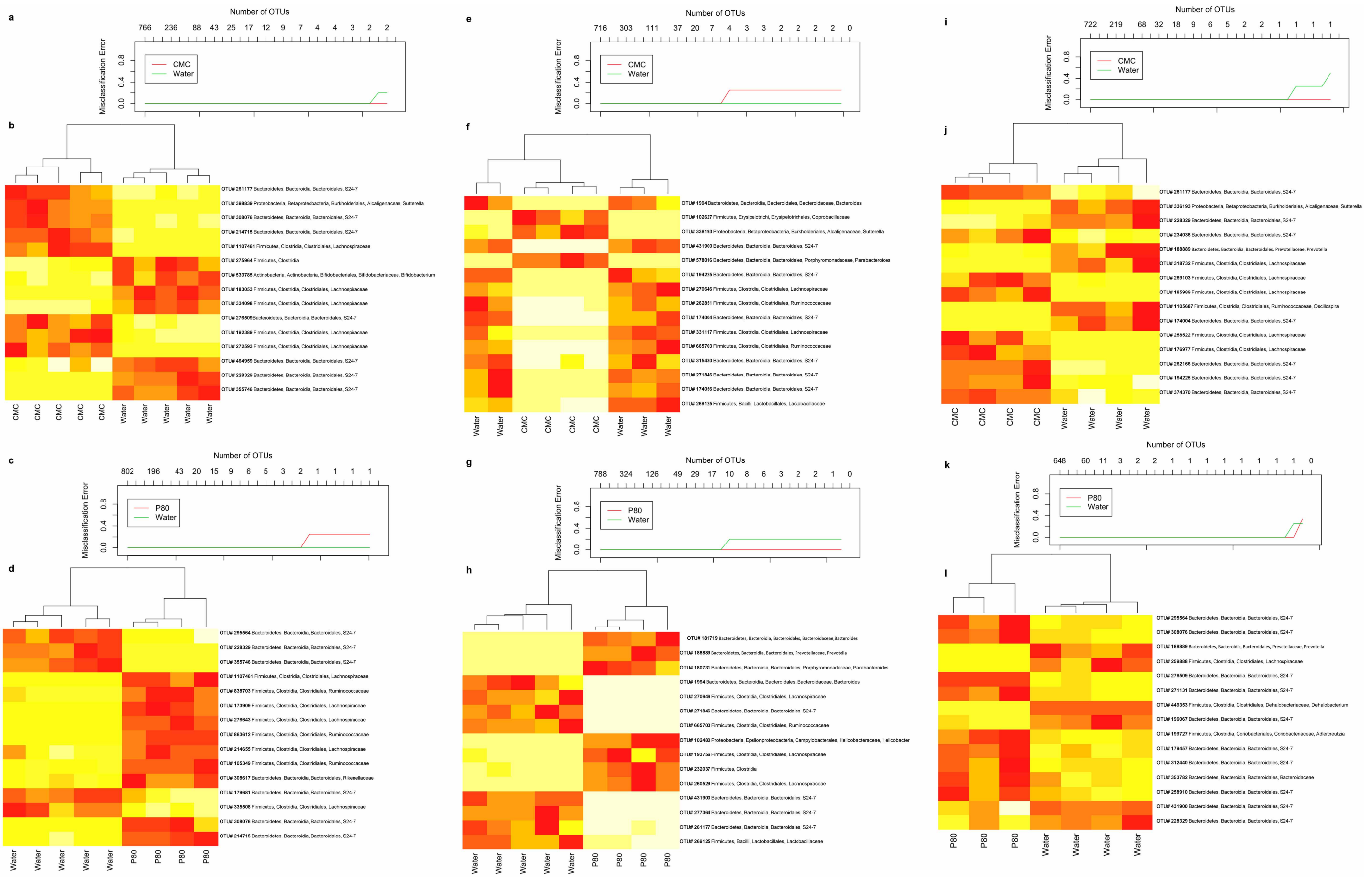
Extended Data Figure 1 | Effects of emulsifiers on mucus-microbiota interaction in wild-type, *Il10*$^{-/-}$ and *Tlr5*$^{-/-}$ mice. **a–d**, Dietary emulsifiers did not affect mucus and mucus-related genes expression in wild-type mice. Wild-type (WT) mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **a–c**, mRNA expression analysis by qRT–PCR of *Muc2* (**a**), *Tff3* (**b**) and *Klf4* (**c**) genes in the colonic mucosa. Points are from individual mice, bar represent the mean ± s.e.m., ($n = 9$). **d**, Colons were stained using periodic acid–Schiff stains. Scale bar, 200 µm. Pictures are representative of 10 biological replicates. **e–j**, Dietary emulsifiers alter microbiota localization, composition and pro-inflammatory potential in *Tlr5*$^{-/-}$ mice. *Tlr5*$^{-/-}$ mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **e–g**, Confocal microscopy analysis of microbiota localization: MUC2, green; actin, purple; bacteria, red; and DNA, blue. Scale bar, 20 µm. Pictures are representative of five biological replicates. **h**, Distances of closest bacteria to intestinal epithelial cells (IEC) per condition over five high-powered fields per mouse. **i**, **j**, PCR-based quantification of total bacterial load (**i**) and bacterial load adhered to colonic mucosa (**j**). Data are the means ± s.e.m., $n = 5$. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test; $*P < 0.05$ compared to water-treated group. **k**, **l**, Dietary emulsifiers do not modify total bacterial load in wild-type and *Il10*$^{-/-}$ mice. Wild-type and *Il10*$^{-/-}$ mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Total bacterial load in stool of wild-type (**k**) and *Il10*$^{-/-}$ (**l**) mice. Points are from individual mice. Data are geometric means with 95% confidence interval ($n = 5$ for **k** except $n = 4$ for CMC- and P80-treated groups; for **l**, $n = 8$, 4 and 6 for water-, CMC- and P80-treated groups, respectively).



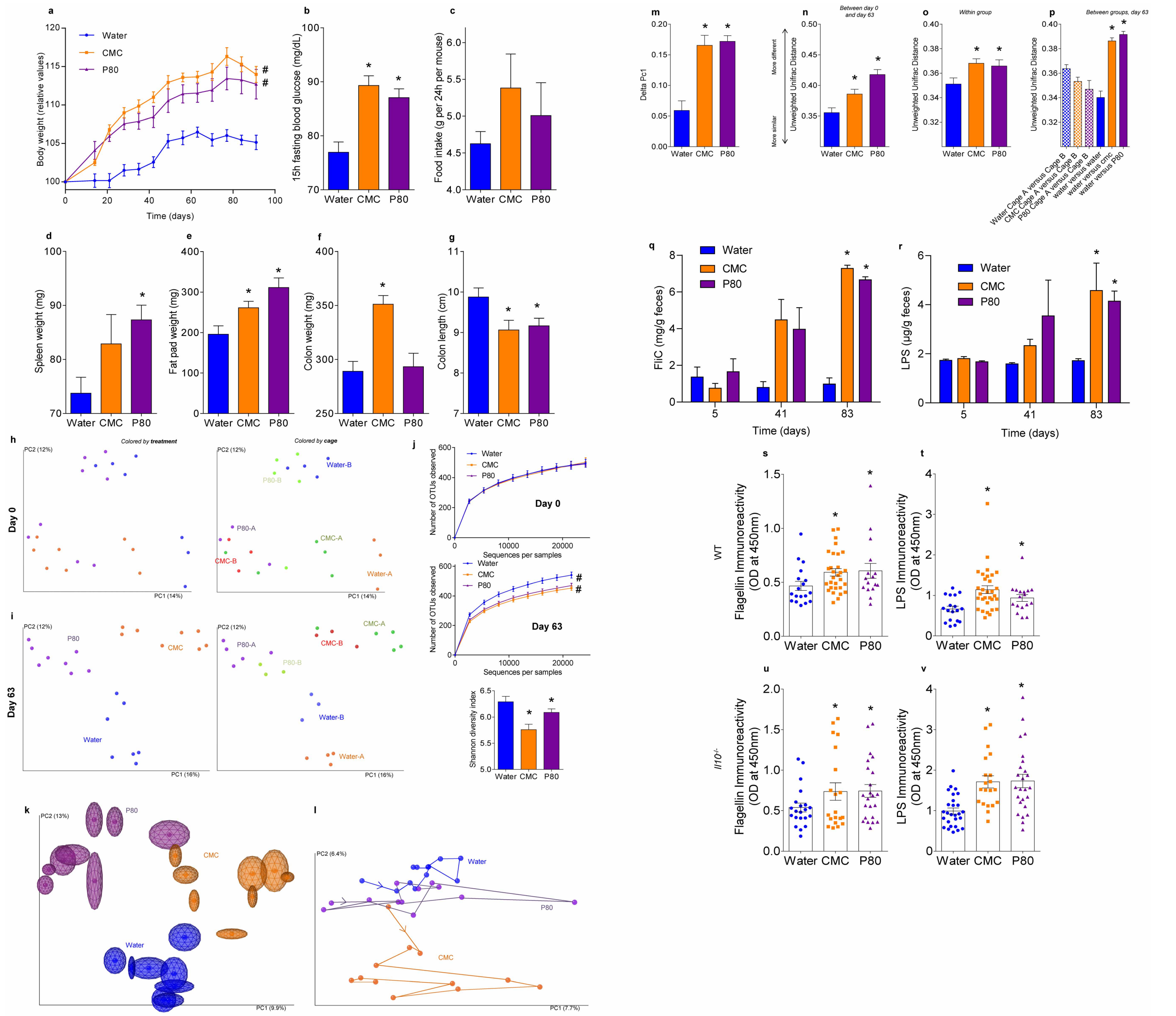
Extended Data Figure 2 | Emulsifiers alter microbiota composition. **a–h**, Dietary emulsifiers induce profound alterations in gut microbiota composition in wild-type and *Il10*$^{-/-}$ mice. Wild-type and *Il10*$^{-/-}$ mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **a**, **b**, **e**, **f**, Day 0 (**a**, **e**) and day 93 (**b**, **f**) microbiota richness and diversity in wild-type (**a**, **b**) and *Il10*$^{-/-}$ (**e**, **f**). **c**, **d**, **g**, **h**, Principal coordinates analysis (PCoA) of the unweighted UniFrac distance matrix of faecal microbiota (**c**, **g**) and mucosa-associated bacteria (**d**, **h**) in wild-type (**c**, **d**) and *Il10*$^{-/-}$ (**g**, **h**) mice. Treatment of each mouse is indicated by point colour and matching coloured circles represent clustering by treatment (blue, water; orange, CMC; purple, P80). Black dashed circles represent mice sharing a cage. Data are the means ± s.e.m.; $n = 5$, except $n = 4$ for P80-treated wild-type mice; $n = 4$ for CMC-treated *Il10*$^{-/-}$ mice; and $n = 9$ for water-treated *Il10*$^{-/-}$ mice. Significance was determined using two-way group ANOVA (#$P < 0.05$) compared to the water-treated group. **i**, **j**, Phylum characterization of emulsifier-induced alteration of gut microbiota composition in wild-type and *Il10*$^{-/-}$ mice. Wild-type and *Il10*$^{-/-}$ mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Relative abundance of phyla are represented for faecal microbiota at day 93 (**i**) and for colonic mucosa-associated bacteria (**j**). Data are the means ± s.e.m., $n = 5$. Significance was determined using two-way ANOVA corrected for multiple comparisons with a Bonferroni test, *$P < 0.05$ compared to water-treated group. **k–o**, Dietary emulsifiers induce profound alterations in gut microbiota composition in *Tlr5*$^{-/-}$ mice. *Tlr5*$^{-/-}$ mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **k**, **l**, Day 0 (**k**) and day 93 (**l**) microbiota richness and diversity ($n = 5$). **m–o**, PCoA of the unweighted UniFrac distance matrix of faecal microbiota at day 0 (**m**), day 93 (**n**) and of mucosa-associated bacteria (**o**). Data are the means ± s.e.m. (for **m**, $n = 4$, 5 and 5 for water-, CMC- and P80-treated groups, respectively; for **n**, $n = 4$, 5 and 3 for water-, CMC- and P80-treated groups, respectively; for **o**, $n = 4$). Significance was determined using two-way group ANOVA (#$P < 0.05$) compared to the water-treated group. **p–t**, Prevalence analysis of OTUs related to mucolytic bacteria in *Il10*$^{-/-}$ mice treated with dietary emulsifier. *Il10*$^{-/-}$ mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. OTUs Prok_MSA # 52947 (**p**; related to *Clostridium perfringens*), 264696 (**q**; related to *Akkermansia muciniphila*), 315982 (**r**; related to *Clostridium perfringens*), 363731 (**s**; related to *Akkermansia muciniphila*), and 178331(**t**; related to *Akkermansia muciniphila*) were analysed. Data are expressed as a percentage of the total sequences analysed and are the means ± s.e.m. ($n = 6$, except $n = 9$ for water-treated *Il10*$^{-/-}$ mice). Significance was determined using two-way ANOVA corrected for multiple comparisons with a Bonferroni test, *$P < 0.05$ compared to water-treated group.



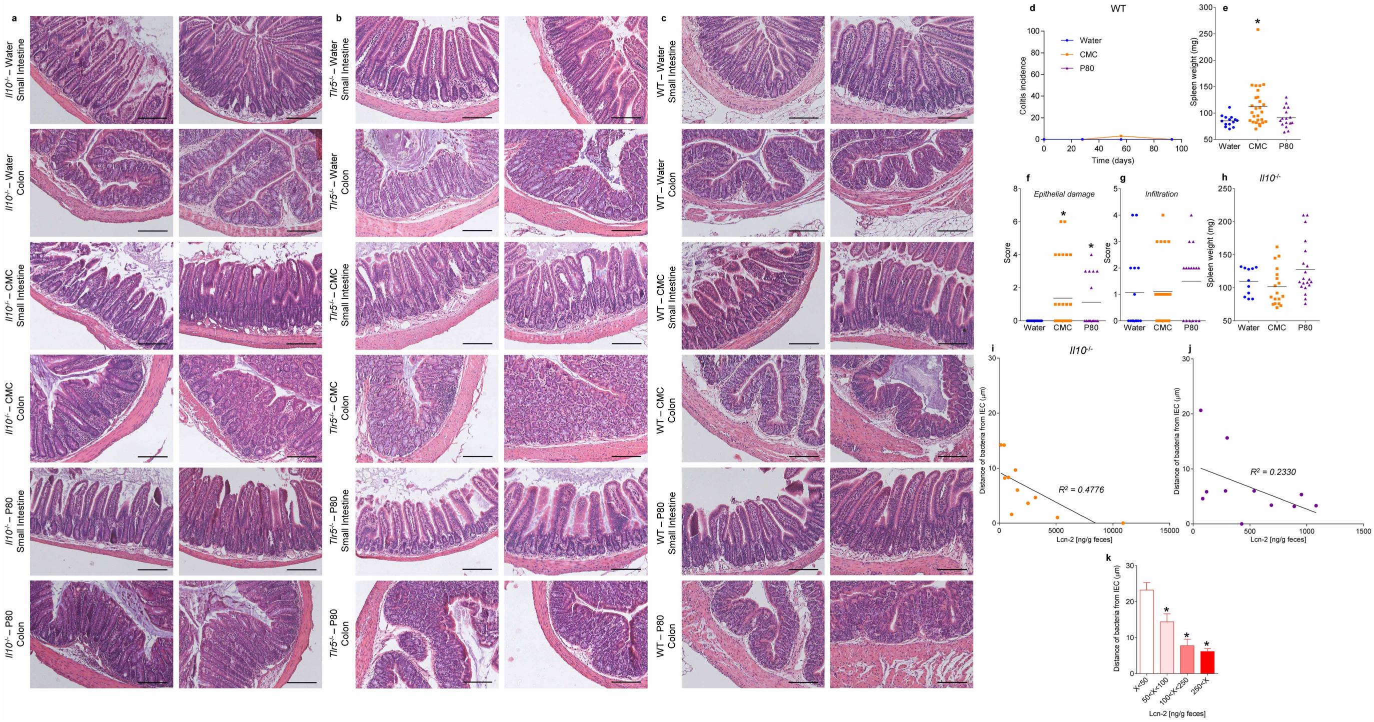
Extended Data Figure 3 | Emulsifier effects on microbiota are irrespective of cage clustering. **a–g**, Dietary emulsifiers alter gut microbiota composition in wild-type mice relative to littermate controls. **a–d**, All the female ($n = 33$; **a**, **b**) and male ($n = 33$; **c**, **d**) mice from 10 different litters were placed into cages in a manner such that each litter was split equally amongst the three experimental groups that were to receive water, CMC or P80 (three cages per sex per condition). Mice were exposed to drinking water containing CMC or P80 (1.0%) for 8 weeks. **a–d**, Day 0 (**a**, **c**) and day 63 (**b**, **d**) Principal coordinates analysis (PCoA) of the unweighted UniFrac distance matrix of faecal microbiota in wild-type mice. Treatment of each mouse is indicated by point colour (blue, water; orange, CMC; purple, P80). The mother of each mouse is indicated by point colour. The cage of each mouse is indicated by point colour. **e**, **f**, Mice were clustered using UPGMA (unweighted pair group method with arithmetic mean). Treatment of each mouse is indicated by line colour (blue, water; orange, CMC; purple, P80). **g**, Schematic representation of the above experimental design. **h–n**, Dietary emulsifiers alter gut microbiota composition in *Tlr5*$^{-/-}$ mice relative to littermate controls. All the female ($n = 25$; **h**, **i**) and male ($n = 23$; **j**, **k**) mice from eight different litters were placed into cages in a manner such that each litter was split equally amongst the three experimental groups that were to receive water, CMC or P80 (two cages per sex per condition). Mice were exposed to drinking water containing CMC or P80 (1.0%) for 8 weeks. **h–k**, Day 0 (**h**, **j**) and day 63 (**i**, **k**) PCoA of the unweighted UniFrac distance matrix of faecal microbiota in *Tlr5*$^{-/-}$ mice. Treatment of each mouse is indicated by point colour (blue, water; orange, CMC; purple, P80). The mother of each mouse is indicated by point colour. The cage of each mouse is indicated by point colour. **l**, **m**, Mice were clustered using UPGMA (unweighted pair group method with arithmetic mean). Treatment of each mouse is indicated by line colour (blue = water; orange = CMC; purple = P80). **n**, Schematic representation of the above experimental design.



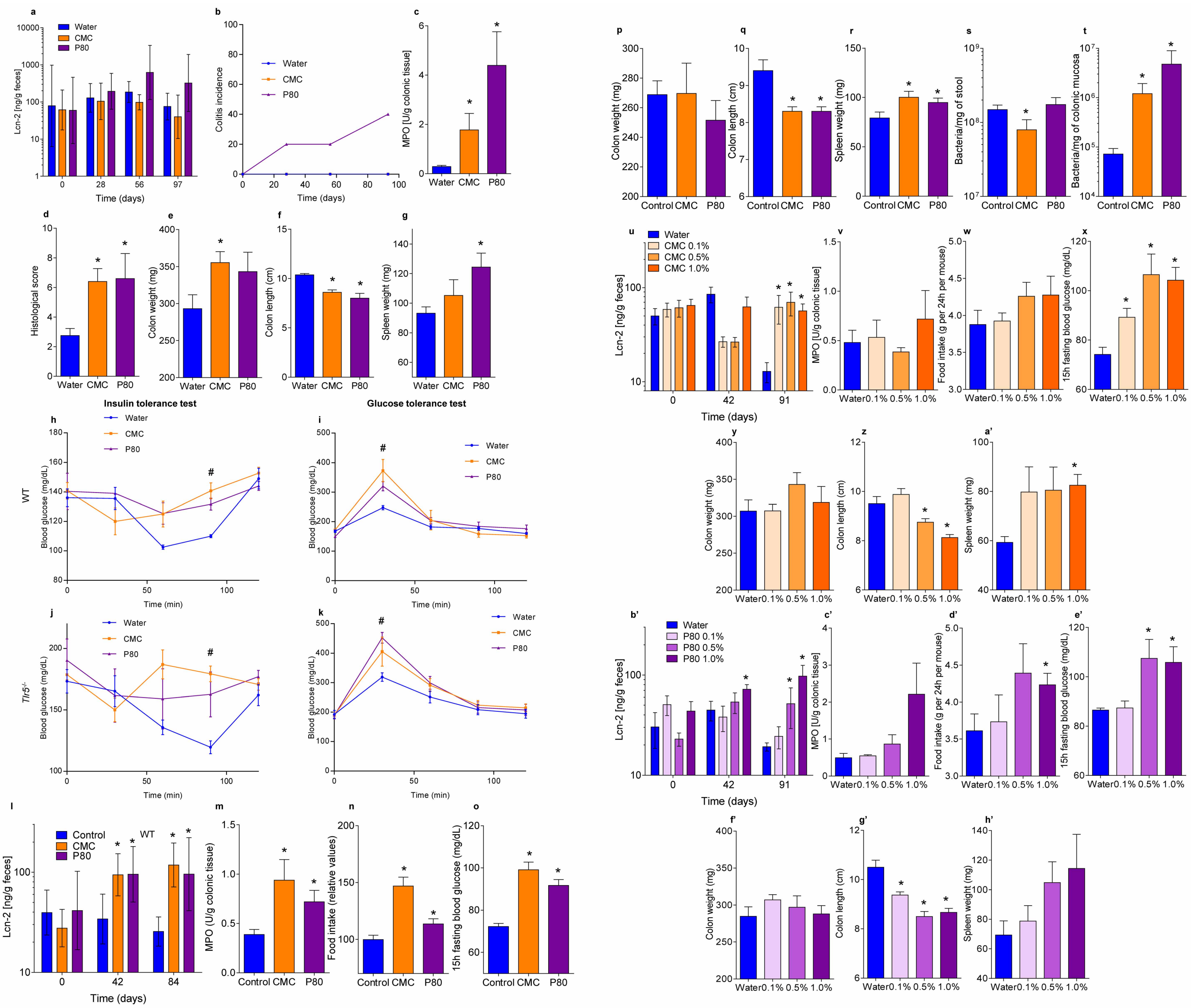
Extended Data Figure 4 | OTUs altered by emulsifiers. a–d, Misclassification error rate and heat map representation of the 15 most significantly altered OTUs in wild-type mice treated with dietary emulsifier. **a–d**, Wild-type mice were exposed to drinking water containing CMC (**a**, **b**) or P80 (**c**, **d**) (1.0%) for 12 weeks. **a**, **c**, Misclassification error rate showing that 15 OTUs were sufficient to successfully discriminate microbiota from each experimental group (error rate = 0). **b**, **d**, Heat map representation of the 15 most significantly altered OTUs in wild-type mice treated with dietary emulsifiers. Colours represent relative expression (white and red for underrepresented and overrepresented, respectively). The 15 OTUs are listed on the right using their Greengenes Prok_MSA identities, and assigned taxonomy are labelled starting phylum, then class, order, family and genus. Dendrogram on the upper part represents sample clustering. **e–h**, As for **a–d** with *Il10*$^{-/-}$ mice. **i–l**, As for **a–d** with *Tlr5*$^{-/-}$ mice. For **a**, **b**, $n = 5$; for **c**, **d**, $n = 5$ and 4 for water- and P80-treated groups, respectively; for **e**, **f**, $n = 5$ and 4 for water- and CMC-treated groups, respectively; for **g**, **h**, $n = 5$ and 4 for water- and P80-treated groups, respectively; for **i**, **j**, $n = 4$; for **k**, **l**, $n = 4$ and 3 for water- and P80-treated groups, respectively.



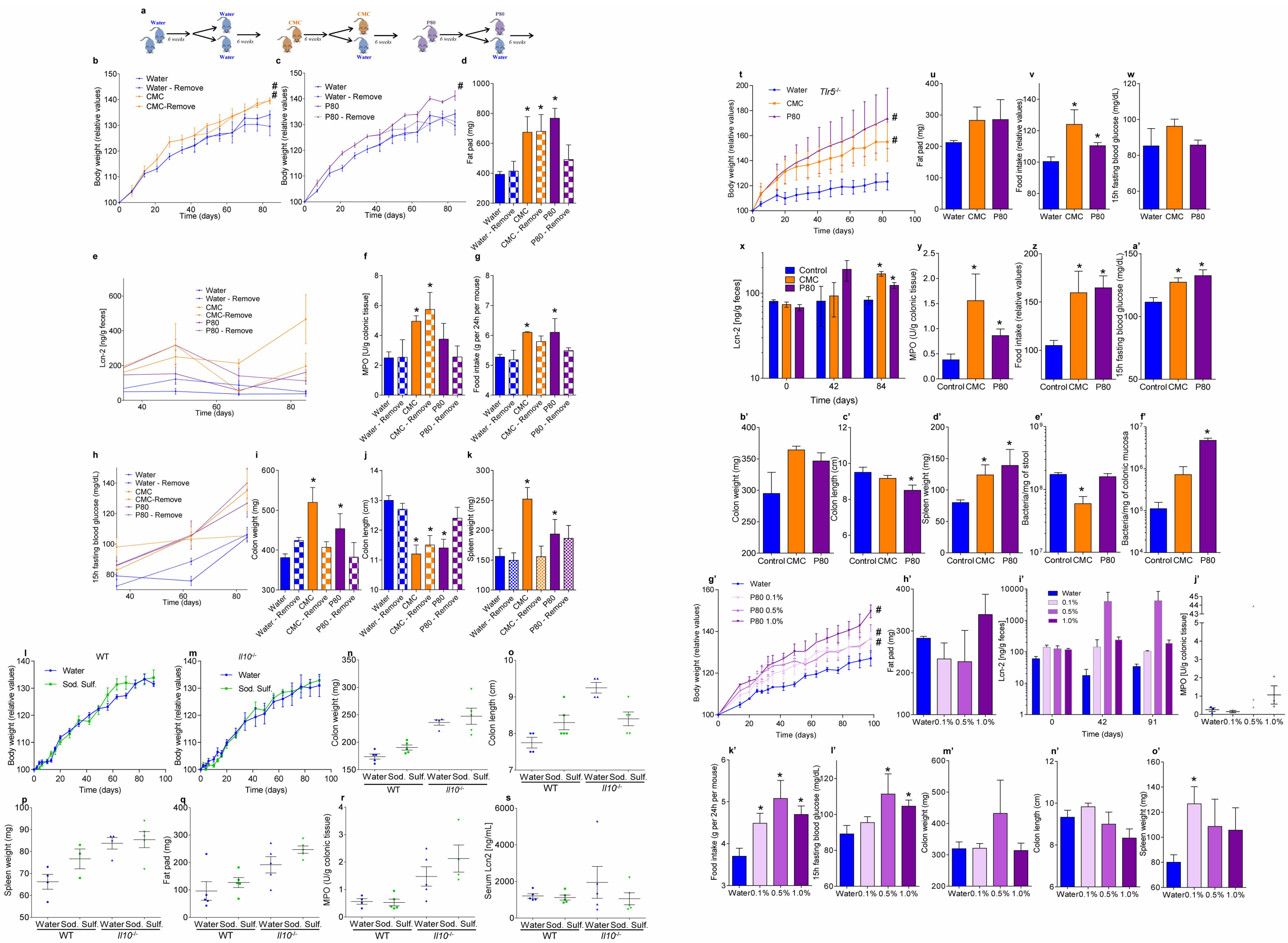
Extended Data Figure 5 | Emulsifier-induced changes in adult mouse microbiota. **a–g**, Dietary emulsifiers promote metabolic syndrome in adult wild-type mice. **a–g**, Four-month-old male wild-type (WT) mice were exposed to drinking water containing CMC or P80 (1.0%) for 8 weeks (two cages per condition). **a**, Body weight over time; **b**, 15 h fasting blood glucose concentration; **c**, food intake measurement; **d**, spleen weights; **e**, fat-pad weights; **f**, colon weights; and **g**, colon lengths. Data are the means ± s.e.m., $n = 10$. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test (*$P < 0.05$ compared to water-treated group) or two-way group ANOVA (#$P < 0.05$ compared to water-treated group). **h–p**, Dietary emulsifiers alter gut microbiota composition in adult wild-type mice. Four-month-old male wild-type mice were exposed to drinking water containing CMC or P80 (1.0%) for 8 weeks (two cages per condition). **h**, **i**, Day 0 (**h**) and day 63 (**i**) Principal coordinates analysis (PCoA) of the unweighted UniFrac distance matrix of faecal microbiota in wild-type mice. Treatment of each mouse is indicated by point colour (blue, water; orange, CMC; purple, P80). The cage of each mouse is indicated by point colour (for **h**, $n = 7$, 8 and 8 for water-, CMC- and P80-treated groups, respectively). **j**, Day 0 and day 63 microbiota richness and diversity. **k**, Day 63 jackknifed PCoA of the unweighted UniFrac distance matrix of faecal microbiota in wild-type mice. Treatment of each mouse is indicated by point colour (blue, water; orange, CMC; purple, P80) ($n = 8$). **l**, After clustering of mouse faecal microbiota using PCoA of the unweighted UniFrac distance matrix, a representative mouse has been used to illustrate the time point evolution of the microbiota ($n = 1$). **m**, After clustering of mouse faecal microbiota using PCoA of the unweighted UniFrac distance matrix, evolution of the principal coordinate 1 between day 0 and day 63 has been calculated for each mouse ($n = 10$). **n**, Average of the UniFrac unweighted distance for each group (water, CMC and P80) between day 0 and day 63 has been calculated ($n = 10$). **o**, Average of the UniFrac unweighted distance within group (water, CMC and P80) has been calculated ($n = 10$). **p**, Average of the UniFrac unweighted distance between group (cages, water, CMC and P80) or within group (water) at day 63 has been calculated ($n = 10$). Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test (*$P < 0.05$ compared to water-treated group) or two-way group ANOVA (#$P < 0.05$ compared to water-treated group). **q**, **r**, Dietary emulsifiers increase pro-inflammatory potential of intestinal microbiota in *Tlr5*$^{-/-}$ mice. *Tlr5*$^{-/-}$ mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Bioactive levels of faecal flagellin (**q**) and LPS (**r**) assayed with TLR5 and TLR4 reporter cells. Data are the means ± s.e.m., $n = 10$. Significance was determined using two-way ANOVA corrected for multiple comparisons with a Bonferroni test; *$P < 0.05$ compared to water-treated group. **s–v**, Dietary emulsifiers increase serum immune reactivity. Wild-type (**s**, **t**) and *Il10*$^{-/-}$ (**u**, **v**) mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Serum immune reactivity (IgG) to flagellin (**s**, **u**) and LPS (**t**, **v**) in wild-type (**s**, **t**) and *Il10*$^{-/-}$ (**u**, **v**) mice. Points are from individual mice. Data are the means ± s.e.m. (for **s**, $n = 18$, 30 and 16 for water-, CMC- and P80-treated groups, respectively; for **t**, $n = 18$, 31 and 17 for water-, CMC- and P80-treated groups, respectively; for **u**, $n = 21$, 20 and 23 for water-, CMC- and P80-treated groups, respectively; for **v**, $n = 27$, 20 and 25 for water-, CMC- and P80-treated groups, respectively). Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test; *$P < 0.05$ compared to water-treated group.



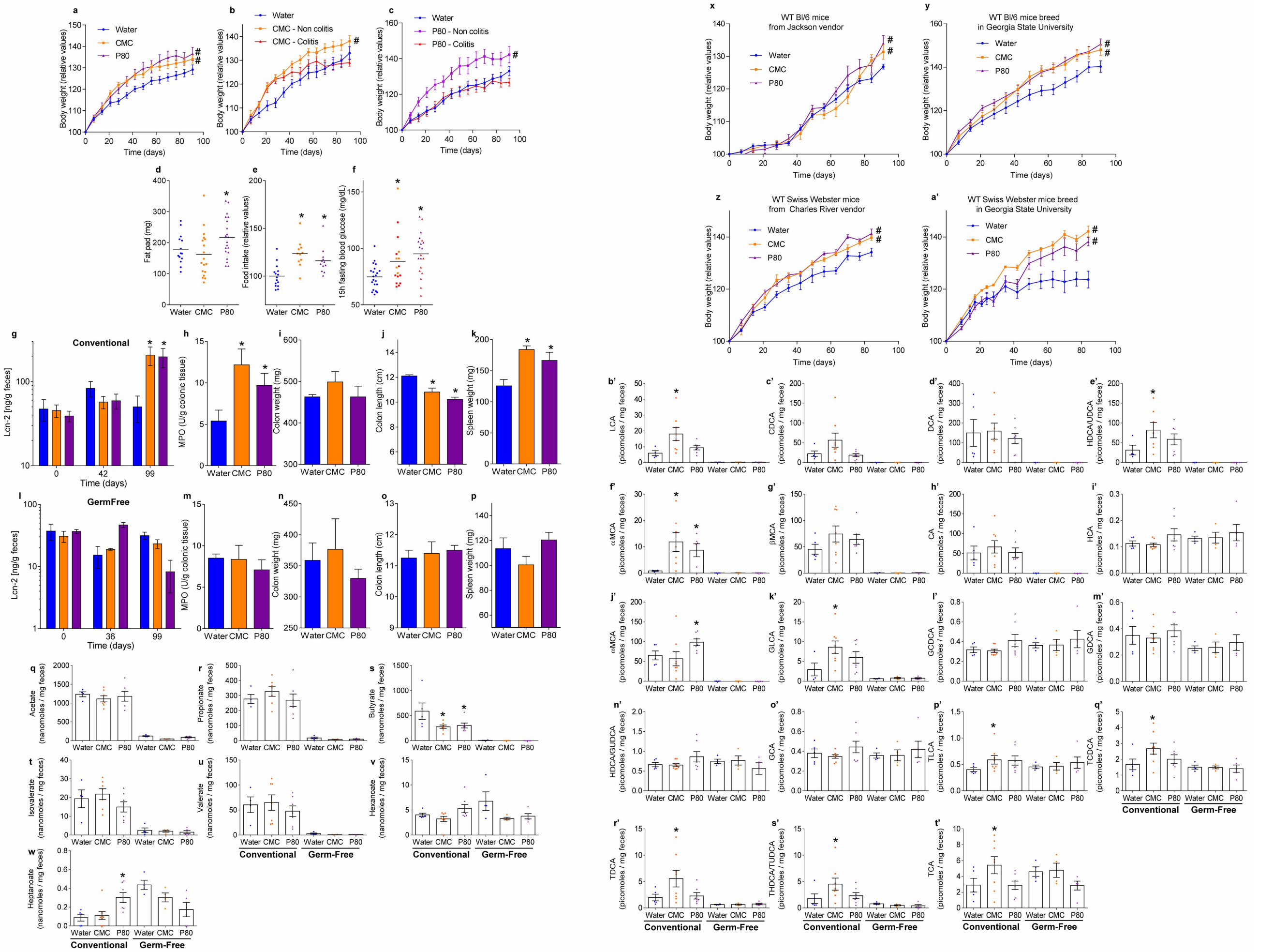
Extended Data Figure 6 | Histopathologic changes in emulsifier-treated wild-type and *Il10*$^{-/-}$ mice. **a**, Dietary emulsifiers induce histopathologically robust inflammation in *Il10*$^{-/-}$ mice. **a**, *Il10*$^{-/-}$ mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Colon and small intestine were haematoxylin and eosin (H&E) stained. Scale bar, 200 µm. Pictures are representatives of 15 biological replicates. **b**, Dietary emulsifiers induce histopathologically robust inflammation in *Tlr5*$^{-/-}$ mice. *Tlr5*$^{-/-}$ mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Colon and small intestine were H&E stained. Scale bar, 200 µm. Pictures are representative of five biological replicates. **c**, Histopathology of emulsifier-treated wild-type (WT) mice. Wild-type mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Colon and small intestine were H&E stained. Scale bar, 200 µm. Pictures are representative of 15 biological replicates. **d**–**h**, Dietary emulsifiers elicit low-grade intestinal inflammation in WT and splenomegaly in *Il10*$^{-/-}$ mice. Wild-type (**d**–**g**) and *Il10*$^{-/-}$ (**h**) mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **d**, Colitis incidence over time; **e**, **h**, spleen weights; **f**, epithelial damage; and **g**, infiltration scores. Points are from individual mice, bars represent the mean. For **e**–**g**, $n = 14$, 27 and 16 for water-, CMC- and P80-treated groups, respectively; for **h**, $n = 11$, 18 and 20 for water-, CMC- and P80-treated groups, respectively. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test; $*P < 0.05$ compared to water-treated group. **i**–**k**, Extent of intestinal inflammation correlates with perturbation in microbiota localization in wild-type and *Il10*$^{-/-}$ mice. *Il10*$^{-/-}$ mice were exposed to drinking water containing CMC (**i**) or P80 (**j**) (1.0%) for 12 weeks. Faecal levels of the inflammatory marker LCN2 as well as confocal microscopy analysis of microbiota localization and estimation of the distances of the closest bacteria to intestinal epithelial cells (IEC) were determined, and plotted in the x and y axis, respectively. Linear regression line was calculated and R^2 was determined; $n = 11$. **k**, Analysis of bacterial–epithelial distance upon stratification of levels of gut inflammatory marker faecal LCN2, using both wild-type and *Il10*$^{-/-}$ mice exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Mice were grouped according to their faecal LCN2 levels and bacterial–epithelial distances were then plotted (mean ± s.e.m.). Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test; $*P < 0.05$ compared to $X < 50$ ng per g group.



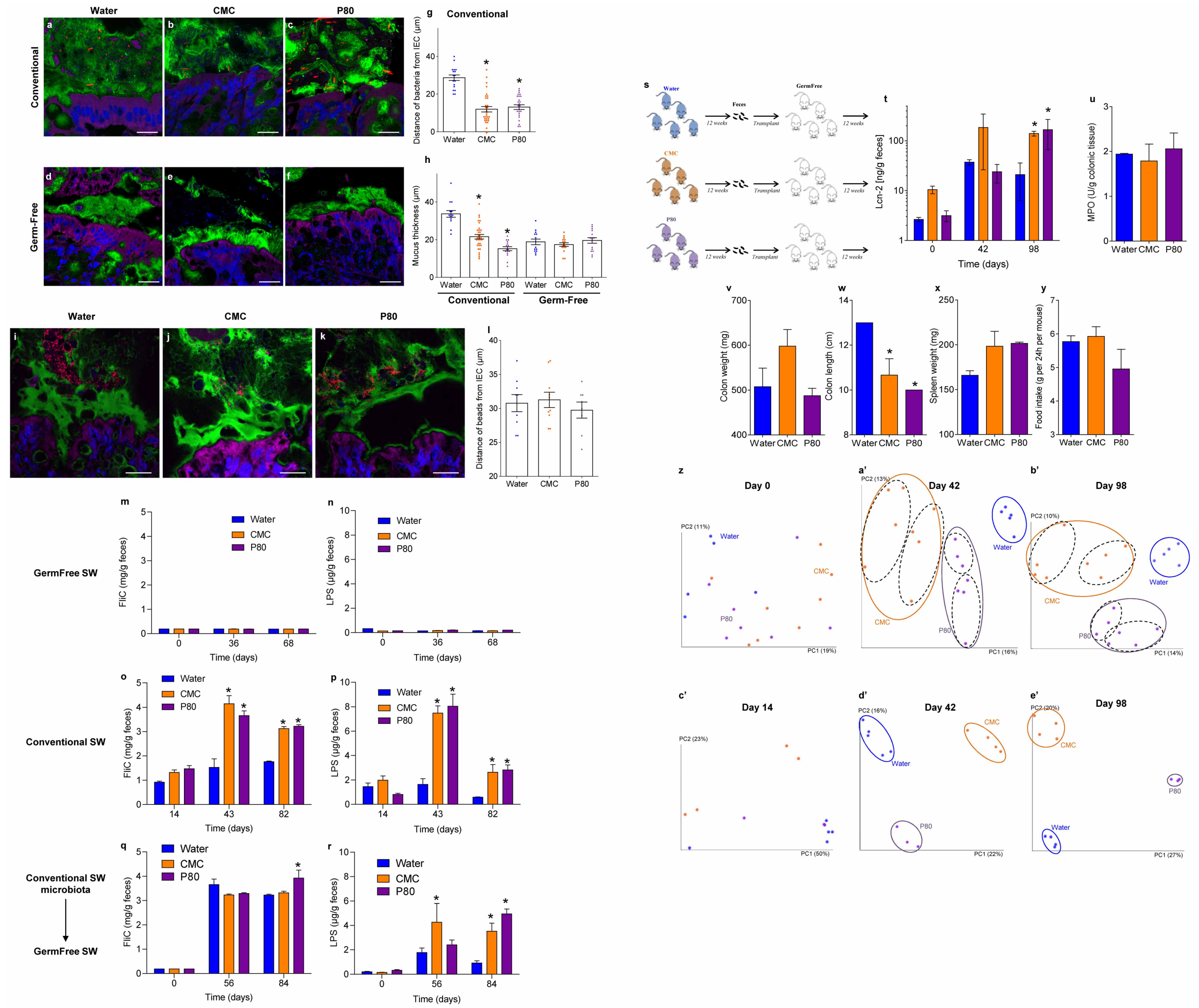
Extended Data Figure 7 | Inflammatory and metabolic parameters in emulsifier-treated wild-type and *Tlr5*$^{-/-}$ mice. **a–g**, Dietary emulsifiers promote intestinal inflammation in *Tlr5*$^{-/-}$ mice. **a–g**, *Tlr5*$^{-/-}$ mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **a**, Faecal levels of the inflammatory marker LCN2 over time; **b**, colitis incidence over time; **c**, myeloperoxidase levels; **d**, histological score; **e**, colon weights; **f**, colon lengths; and **g**, spleen weights. Data are the means ± s.e.m. or geometric means with 95% confidence interval (the latter for **a**), $n = 5$. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test or using two-way ANOVA corrected for multiple comparisons with a Bonferroni test; $*P < 0.05$ compared to water-treated group. **h–k**, Dietary emulsifiers induce metabolic syndrome in wild-type (WT) and *Tlr5*$^{-/-}$ mice. Wild-type (**h**, **i**) and *Tlr5*$^{-/-}$ (**j**, **k**) mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Glucose tolerance (**h**, **j**) and insulin sensitivity (**i**, **k**) were analysed. Data are the means ± s.e.m., $n = 5$. Significance was determined using two-way group ANOVA ($\#P < 0.05$) compared to water-treated group. **l–t**, Emulsifier-supplemented chow elicits low-grade intestinal inflammation in wild-type mice. Wild-type mice were given mouse chow containing CMC or P80 (1.0%) for 12 weeks. **l**, Faecal levels of the inflammatory marker LCN2 over time; **m**, myeloperoxidase levels; **n**, food intake measurement; **o**, 15 h fasting blood glucose concentration; **p**, colon weights; **q**, colon lengths; **r**, spleen weights; and **s**, **t**, PCR-based quantification of total bacterial load (**s**) and bacterial load adhered to colonic mucosa (**t**). Data are the means ± s.e.m. or geometric means with 95% confidence interval (the latter for **l**), $n = 5$. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test or using two-way ANOVA corrected for multiple comparisons with a Bonferroni test; $*P < 0.05$ compared to control group. **u–h′** Dose-response characterization of dietary emulsifiers on intestinal inflammation. Wild-type mice were exposed to drinking water containing 0.1–1.0% CMC (**u–a′**) or P80 (**b′–h′**) for 12 weeks. **u**, **b′**, Faecal levels of the inflammatory marker LCN2 over time; **v**, **c′**, myeloperoxidase levels; **w**, **d′**, food intake measurement; **x**, **e′**, 15 h fasting blood glucose concentration; **y**, **f′**, colon weights; **z**, **g′**, colon lengths; and **a′**, **h′**, spleen weights. Data are the means ± s.e.m., $n = 5$. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test or using two-way ANOVA corrected for multiple comparisons with a Bonferroni test; $*P < 0.05$ compared to water-treated group.



Extended Data Figure 8 | Reversibility and dose dependence of emulsifier-induced effects on inflammation and metabolism. **a–k**, Emulsifier-induced metabolic syndrome in Swiss Webster mice is partially reversible by 6 weeks after emulsifier treatment. **a**, Schematic representation of the experiment. **b**, **c**, Body weight over time; **d**, fat-pad weight; **e**, faecal levels of the inflammatory marker LCN2 over time; **f**, myeloperoxidase levels; **g**, food intake measurement; **h**, 15 h fasting blood glucose concentration; **i**, colon weights; **j**, colon lengths; and **k**, spleen weights. Data are the means ± s.e.m., $n = 5$. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test (*$P < 0.05$ compared to water-treated group) or two-way group ANOVA (#$P < 0.05$ compared to water-treated group). **l–s**, Sodium sulfite did not induce robust or low-grade intestinal inflammation. Wild-type and *Il10*$^{-/-}$ mice were exposed to drinking water containing sodium sulfite (1.0%) for 12 weeks. **l**, **m**, Body weight over time; **n**, colon weights; **o**, colon lengths; **p**, spleen weights; **q**, fat-pad weight; **r**, myeloperoxidase levels; and **s**, serum levels of the inflammatory marker LCN2. Data are the means ± s.e.m., $n = 5$. Points are from individual mice. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test. **t–w**, Dietary emulsifiers promote metabolic syndrome in *Tlr5*$^{-/-}$ mice. *Tlr5*$^{-/-}$ mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **t**, Body weight over time; **u**, fat-pad weight; **v**, food intake measurement; and **w**, 15 h fasting blood glucose concentration. Data are the means ± s.e.m., $n = 5$. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test (*$P < 0.05$ compared to water-treated group) or two-way group ANOVA (#$P < 0.05$ compared to water-treated group). **x–f′**, Emulsifier-supplemented chow promotes intestinal inflammation in *Tlr5*$^{-/-}$ mice. *Tlr5*$^{-/-}$ mice were given mouse chow containing CMC or P80 (1.0%) for 12 weeks. **x**, Faecal levels of the inflammatory marker LCN2 over time; **y**, myeloperoxidase levels; **z**, food intake measurement; **a′**, 15 h fasting blood glucose concentration; **b′**, colon weights; **c′**, colon lengths; **d′**, spleen weights; and **e′**, **f′**, PCR-based quantification of total bacterial load (**e′**) and bacterial load adhered to colonic mucosa (**f′**). Data are the means ± s.e.m., $n = 5$. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test or using two-way ANOVA corrected for multiple comparisons with a Bonferroni test; *$P < 0.05$ compared to control group. **g′–o′**, Dose-response characterization of dietary emulsifiers on intestinal inflammation in *Tlr5*$^{-/-}$ mice. *Tlr5*$^{-/-}$ mice were exposed to drinking water containing 0.1–1.0% P80 for 12 weeks. **g′**, Body weight over time; **h′**, fat-pad weight; **i′**, faecal levels of the inflammatory marker LCN2 over time; **j′**, myeloperoxidase levels; **k′**, food intake measurement; **l′**, 15 h fasting blood glucose concentration; **m′**, colon weights; **n′**, colon lengths; and **o′**, spleen weights. Data are the means ± s.e.m., $n = 3$. Points in **d** are from individual mice. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test (*$P < 0.05$ compared to water-treated group) or two-way group ANOVA (#$P < 0.05$ compared to water-treated group).



Extended Data Figure 9 | Effects of emulsifiers on metabolic parameters and bile acids in conventional and germ-free mice. **a**–**f**, Dietary emulsifiers promotes metabolic syndrome in *Il10*$^{-/-}$ mice. *Il10*$^{-/-}$ mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **a**–**c**, Body weight over time; **d**, fat-pad weight; **e**, food intake measurement; and **f**, 15 h fasting blood glucose concentration. Data are the means ± s.e.m. (for **a**, $n = 24$, 18 and 21 for water-, CMC- and P80-treated groups, respectively; for **b**, $n = 14$, 11 and 8 for water-, CMC- and P80-treated groups, respectively; for **c**, $n = 14$, 9 and 9 for water-, CMC- and P80-treated groups, respectively; for **d**, $n = 14$, 18 and 20 for water-, CMC- and P80-treated groups, respectively; for **e**, $n = 15$, 11 and 12 for water-, CMC- and P80-treated groups, respectively; for **f**, $n = 21$, 17 and 20 for water-, CMC- and P80-treated groups, respectively). Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test (*$P < 0.05$ compared to water-treated group) or two-way group ANOVA (#$P < 0.05$ compared to water-treated group). Points are from individual mice and red points in **f** represent mice with overt colitis. **g**–**p**, Emulsifier-induced low-grade intestinal inflammation was abolished under germ-free conditions. Conventionally housed (**g**–**k**) and germ-free (**l**–**p**) Swiss Webster mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **g**, **l**, Faecal levels of the inflammatory marker LCN2 over time; **h**, **m**, myeloperoxidase levels; **i**, **n**, colon weights; **j**, **o**, colon lengths; and **k**, **p**, spleen weights. Data are the means ± s.e.m.; $n = 8$ for conventionally housed mice and $n = 4$ for germ-free mice. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test or using two-way ANOVA corrected for multiple comparisons with a Bonferroni test, *$P < 0.05$ compared to control group. **q**–**w**, Dietary emulsifiers induce perturbations in faecal short-chain fatty acid composition. Faecal short-chain fatty acids composition was analysed at the Metabolomics Core of the University of Michigan. Acetate (**q**), propionate (**r**), butyrate (**s**), isovalerate (**t**), valerate (**u**), hexanoate (**v**) and heptanoate (**w**) were analysed. Data are the means ± s.e.m. ($n = 5$, 8 and 7 for water-, CMC- and P80-treated conventional mice groups, respectively; $n = 4$, 4 and 5 for water-, CMC- and P80-treated germ-free mice groups, respectively). Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test; *$P < 0.05$ compared to water-treated group. **x**–**a′**, Dietary emulsifiers promote metabolic syndrome in mice from different vendors. Wild-type mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **x**, **y**, Body weight over time of Bl/6 mice used upon receipt from Jackson Laboratories (**x**) or bred at Georgia State University (**y**). **z**, **a′**, Body weight over time of Swiss Webster mice used upon receipt from Charles River company (**z**) or bred at Georgia State University (**a′**). Data in **x** are not used elsewhere in report, while data in **y**, **z**, and **a′** are from Figs 3a and 4a and Extended Data Fig 8a–k. Data are the means ± s.e.m. $n = 8$ for Bl/6 mice used upon receipt from Jackson Laboratories, $n = 16$ for Bl/6 mice bred at Georgia State University, $n = 10$ for Swiss Webster mice used upon receipt from Charles River company, $n = 8$ for Swiss Webster mice bred at Georgia State University. Significance was determined using two-way group ANOVA (#$P < 0.05$ compared to water-treated group). **b′**–**t′**, Dietary emulsifiers induce perturbations in faecal bile acids composition. Faecal bile acids composition was analysed at the Metabolomics Core of the University of Michigan. Lithocholic acid (LCA; **b′**), chenodeoxycholic acid (CDCA; **c′**), deoxycholic acid (DCA; **d′**), hyodeoxycholic acid/ursodeoxycholic acid (HDCA/UDCA; **e′**), α-muricholic acid (α-MCA; **f′**), β-muricholic acid (β -MCA; **g′**), cholic acid (CA; **h′**), hyocholic acid (HCA; **i′**), ω-muricholic acid (ω-MCA; **j′**), glycolithocholic acid (GLCA; **k′**), glycochenodeoxycholic acid (GCDCA; **l′**), glycodeoxycholic acid (GDCA; **m′**), hyodeoxycholic acid/glycoursodeoxycholic acid (HDCA/GUDCA; **n′**), glycocholic acid (GCA; **o′**), taurolithocholic acid (TLCA; **p′**), taurine-conjugated chenodeoxycholic acid (TCDCA; **q′**), taurodeoxycholic acid/tauroursodeoxycholic acid (TDCA/TUDCA; **r′**), taurohyodeoxycholic acid (**s′**), and taurocholic acid (TCA; **t′**) were analysed. Data are the means ± s.e.m. ($n = 5$, 8 and 7 for water-, CMC- and P80-treated conventional mice groups, respectively; $n = 4$, 4 and 5 for water-, CMC- and P80-treated germ-free mice groups, respectively). Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test; *$P < 0.05$ compared to water-treated group.



Extended Data Figure 10 | Faecal transplants transfer some effects of emulsifers. a–h, Dietary emulsifiers do not alter mucus thickness under germ-free conditions. **a–h**, Conventionally-housed (**a–c**, **g–h**) and germ-free Swiss Webster (SW; **d–f**, **h**) mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. **a–f**, Confocal microscopy analysis of microbiota localization: MUC2, green; actin, purple; bacteria, red; and DNA, blue. Scale bar, 20 µm. **g**, Distances of closest bacteria to intestinal epithelial cells (IEC) per condition over five high-powered fields per mouse. Pictures are representative of five biological replicates. **h**, Mucus thickness over five high-powered fields per mouse. Data are the means ± s.e.m., $n = 5$. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test; $*P < 0.05$ compared to water-treated group. **i–l**, Dietary emulsifiers do not induce drastic perturbations of mucus layer integrity under germ-free conditions. Germ-free Swiss Webster mice were exposed to drinking water containing CMC or P80 (1.0%) for 8 weeks. Mice were removed from the isolator, inoculated with 0.5 µm green fluorescent beads (Polysciences, Warrington, Pennsylvania), and euthanized 7 h post-inoculation. **i–k**, Confocal microscopy analysis of fluorescent beads localization: MUC2, green; actin, purple; fluorescent beads, red; and DNA, blue. Scale bar, 20 µm. **l**, Distances of closest fluorescent beads to intestinal epithelial cells per condition over five high-powered fields per mouse. Pictures are representatives of five biological replicates. Data are the means ± s.e.m., $n = 4$. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test. **m–r**, Dietary emulsifiers increase pro-inflammatory potential of intestinal microbiota in Swiss Webster mice, transferable to germ-free mice recipients. **m, n**, Germ-free Swiss Webster mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Bioactive levels of faecal flagellin (**m**) and LPS (**n**) were assayed with TLR5 and TLR4 reporter cells. **o, p**, Swiss Webster mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Bioactive levels of faecal flagellin (**o**) and LPS (**p**) were assayed with TLR5 and TLR4 reporter cells. **q, r**, Germ-free Swiss Webster mice were conventionalized via microbiota transplant from the Swiss Webster mice treated with emulsifiers described above. Bioactive levels of faecal flagellin (**q**) and LPS (**r**) were assayed with TLR5 and TLR4 reporter cells. Data are the means ± s.e.m. $n = 4$ for germ-free mice, $n = 8$ for conventionally housed mice and $n = 5$ for conventionalized mice. Significance was determined using two-way ANOVA corrected for multiple comparisons with a Bonferroni test; $*P < 0.05$ compared to control group. **s–y**, Microbiota transplant transfers emulsifier-induced low-grade intestinal inflammation. Germ-free Swiss Webster mice were conventionalized via microbiota transplant from mice that received standard drinking water or drinking water containing CMC or P80 (1.0%). **s**, Schematic representation of the experiment. **t**, Faecal levels of the inflammatory marker LCN2 over time; **u**, myeloperoxidase levels; **v**, colon weights; **w**, colon lengths; **x**, spleen weights; and **y**, food intake measurement. Data are the means ± s.e.m., $n = 4$. Significance was determined using one-way ANOVA corrected for multiple comparisons with a Sidak test or using two-way ANOVA corrected for multiple comparisons with a Bonferroni test; $*P < 0.05$ compared to control group. **z–e′**, Dietary emulsifiers induce profound alterations in gut microbiota composition in Swiss Webster mice, transferable to germ-free mice recipients. **z–b′**, Swiss Webster mice were exposed to drinking water containing CMC or P80 (1.0%) for 12 weeks. Principal coordinates analysis (PCoA) of the unweighted UniFrac distance matrix of faecal microbiota at day 0 (**z**), day 42 (**a′**) and day 98 (**b′**). **c′–e′**, Germ-free Swiss Webster mice were conventionalized via microbiota transplant from the Swiss Webster mice treated with emulsifiers described above. PCoA of the unweighted UniFrac distances of faecal microbiota at day 14 (**c′**), day 42 (**d′**) and day 98 (**e′**) post-transplant. For **z–b′**, $n = 5$, 8 and 7 for water-, CMC- and P80-treated groups, respectively; for **c′–e′**, $n = 5$, 4 and 3 for water-, CMC- and P80-treated groups, respectively. Treatment of each mouse is indicated by point colour and matching coloured circles indicate mice receiving the same treatment (blue, water; orange, CMC; purple, P80). Black dashed circles represent mice sharing a cage.

LETTER

doi:10.1038/nature14021

Dauer-independent insulin/IGF-1-signalling implicates collagen remodelling in longevity

Collin Y. Ewald[1,2,3], Jess N. Landis[4]*, Jess Porter Abate[1,2,3]*, Coleen T. Murphy[4] & T. Keith Blackwell[1,2,3]

Interventions that delay ageing mobilize mechanisms that protect and repair cellular components[1–3], but it is unknown how these interventions might slow the functional decline of extracellular matrices[4,5], which are also damaged during ageing[6,7]. Reduced insulin/IGF-1 signalling (rIIS) extends lifespan across the evolutionary spectrum, and in juvenile *Caenorhabditis elegans* also allows the transcription factor DAF-16/FOXO to induce development into dauer, a diapause that withstands harsh conditions[1,2]. It has been suggested that rIIS delays *C. elegans* ageing through activation of dauer-related processes during adulthood[2,8,9], but some rIIS conditions confer robust lifespan extension unaccompanied by any dauer-like traits[1,10,11]. Here we show that rIIS can promote *C. elegans* longevity through a program that is genetically distinct from the dauer pathway, and requires the Nrf (NF-E2-related factor) orthologue SKN-1 acting in parallel to DAF-16. SKN-1 is inhibited by IIS and has been broadly implicated in longevity[12–14], but is rendered dispensable for rIIS lifespan extension by even mild activity of dauer-related processes. When IIS is decreased under conditions that do not induce dauer traits, SKN-1 most prominently increases expression of collagens and other extracellular matrix genes. Diverse genetic, nutritional, and pharmacological pro-longevity interventions delay an age-related decline in collagen expression. These collagens mediate adulthood extracellular matrix remodelling, and are needed for ageing to be delayed by interventions that do not involve dauer traits. By genetically delineating a dauer-independent rIIS ageing pathway, our results show that IIS controls a broad set of protective mechanisms during *C. elegans* adulthood, and may facilitate elucidation of processes of general importance for longevity. The importance of collagen production in diverse anti-ageing interventions implies that extracellular matrix remodelling is a generally essential signature of longevity assurance, and that agents promoting extracellular matrix youthfulness may have systemic benefit.

We hypothesized that SKN-1 would be required for rIIS lifespan extension under conditions in which dauer-associated processes are inactive. Class 2 mutations in the insulin/IGF-1 receptor DAF-2 induce adulthood dauer-related traits that are mild at 20 °C, and severe at 22.5 °C or above, but class 1 mutations do not (Supplementary Videos 1 and 2 and Supplementary Discussion)[10]. SKN-1 is inhibited by IIS phosphorylation but is dispensable for dauer development[13], adulthood dauer-related traits (Extended Data Fig. 1a–d and Supplementary Table 1), or lifespan extension by Class 2 *daf-2* mutations at 20 °C (Extended Data Fig. 1a and Supplementary Table 2)[13]. By contrast, at 15 °C SKN-1 was completely required for longevity in the same class 2 *daf-2* mutants (Fig. 1a, Extended Data Fig. 1a, e, Extended Data Table 1 and Supplementary Table 2), which do not show dauer traits at 15 °C (ref. 10) because low temperature inhibits dauer entry (Supplementary Discussion). The *skn-1* gene was also essential at 20 °C in class 2 *daf-16; daf-2* double mutants that expressed DAF-16 specifically in the intestine, a condition that rescues longevity but not dauer development[1,15] or traits (Extended Data Fig. 1f, g and Extended Data Table 1). Finally, *skn-1* was required at 15, 20, or 25 °C for lifespan extension from *daf-2* RNA interference (RNAi) (Fig. 1b, Extended Data Fig. 1a, Extended Data Table 1 and Supplementary Table 2), which promotes dauer entry only at extreme temperature and does not induce dauer traits in adults (Extended Data Fig. 1h–j). In these last two scenarios, the absence of dauer traits may reflect DAF-16 insufficiency in neurons, which are central to dauer regulation[15,16] and resistant to RNAi (Extended Data Fig. 1h, i and Extended Data Table 1). Lifespan extension is extremely robust when *daf-2* RNAi is performed in the class 1 mutant *daf-2*(*e1368*)[11], which lacks adulthood dauer traits but predisposes to dauer entry[10]. The *skn-1* gene was largely required for this lifespan extension at 20 °C, and was essential for the even greater healthy lifespan extension seen at 15 °C (117 days maximum; Fig. 1c, d, Extended Data Fig. 1a and Extended Data Table 1).

The *skn-1* dependence of rIIS longevity tracked inversely with predisposition to dauer entry or adulthood dauer traits, and was not determined

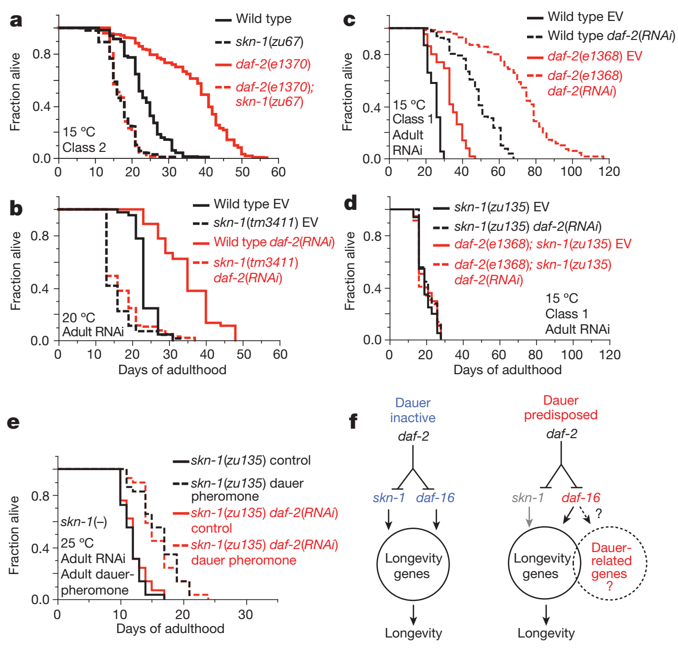


Figure 1 | Dauer-independent rIIS longevity requires SKN-1. a, **b**, The *skn-1* dependence of rIIS lifespan extension in the absence of dauer traits. **c**, **d**, The *skn-1* dependence of extreme rIIS longevity. EV, empty RNAi vector. **e**, The *skn-1* independence of longevity from adulthood dauer pheromone treatment but not *daf-2*(*RNAi*). **f**, Longevity assurance programs regulated by IIS. Under conditions that predispose to dauer traits (right), some SKN-1 functions may be assumed by DAF-16, possibly including ECM remodelling. Statistics and additional lifespan data are in Extended Data Table 1 and Supplementary Table 2.

[1]Joslin Diabetes Center, One Joslin Place, Boston, Massachusetts 02215, USA. [2]Harvard Stem Cell Institute, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA. [3]Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02215, USA. [4]Department of Molecular Biology, Lewis-Sigler Institute for Integrative Genomics, Princeton University, 148 Carl Icahn Laboratory, Washington Road, Princeton, New Jersey 08544, USA.
*These authors contributed equally to this work.

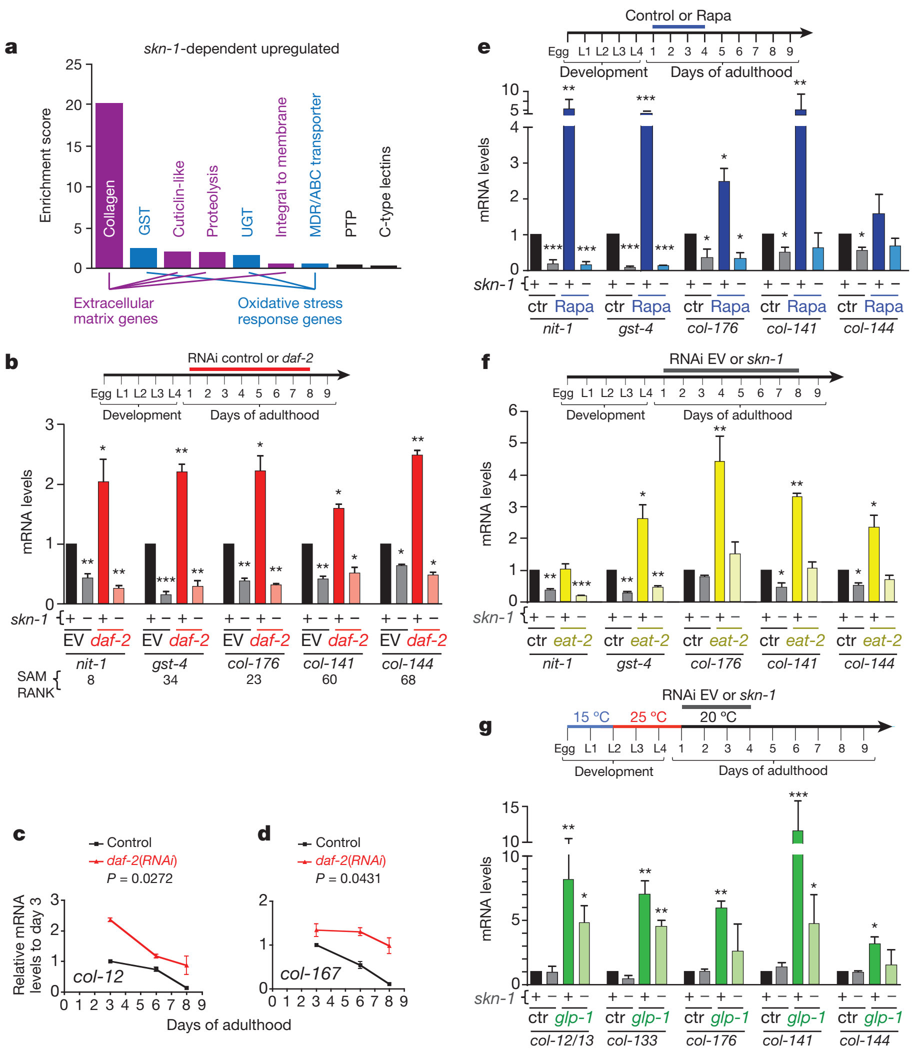


Figure 2 | Longevity-promoting interventions increase *skn-1*-dependent collagen expression in adults. **a**, Functional categories enriched in SKN-1-upregulated *daf-2*(−) gene sets, identified by the Database for Annotation, Visualization, and Integrated Discovery (DAVID). Enrichment scores of at least 1.3 are shown. **b**–**d**, Collagen upregulation by adulthood *daf-2* RNAi. Messenger RNA (mRNA) expression in wild-type (+) or *skn-1*(*zu135*) (−) animals, assayed by quantitative PCR (qPCR). The *nit-1* and *gst-4* genes are canonical SKN-1 targets[14,18]. Significance analysis of microarray (SAM) score ranks are in Supplementary Table 3. **e**, Rapamycin-treated (100 µM) wild-type and *skn-1*(*zu67*) animals are compared. **f**, Expression in the dietary restriction model *eat-2*. RNAi-sensitized control (*rrf-3*(*pk1426*)) (ctr) or *eat-2*(*ad1116*); *rrf-3*(*pk1426*) (*eat-2*) adults were exposed to empty RNAi vector or *skn-1* RNAi. **g**, Upregulation after germline stem-cell proliferation block induced by *glp-1*(*bn18*) temperature shift. Three replicates of 200 worms were analysed at the indicated days (**c**, **d**) or at the end of treatment. Data are mean ± s.e.m. $*P < 0.05$, $**P < 0.001$, $***P < 0.0001$ relative to wild type or control, by one-sample *t*-test, two-tailed, hypothetical mean of 1.

by temperature (Extended Data Fig. 1a). Also, *skn-1* dependence did not correlate with the magnitude of rIIS lifespan extension, suggesting that it was not determined by the extent of IIS reduction (Extended Data Fig. 1a). Accordingly, DAF-16 and SKN-1 nuclear localization was increased as robustly by *daf-2* RNAi as by class 1 or class 2 *daf-2* mutations, and was similar in *daf-2* mutants at 15 and 20 °C (Extended Data Fig. 1k–o). Activation of dauer processes in adults by a mechanism other than genetic IIS reduction should extend lifespan without *skn-1*. Accordingly, *skn-1* was dispensable for lifespan extension from adulthood dauer pheromone exposure (Fig. 1e, Extended Data Fig. 1p, q and Extended Data Table 1).

We conclude that *skn-1* is needed for rIIS longevity specifically when dauer-associated mechanisms are inactive (Extended Data Fig. 1a). This genetic requirement for *skn-1* reveals that rIIS extends lifespan through two downstream pathways that may overlap (Fig. 1f). During the reproductive life cycle, IIS inhibits a protective program that requires both DAF-16 and SKN-1, and does not involve dauer-specific processes. This

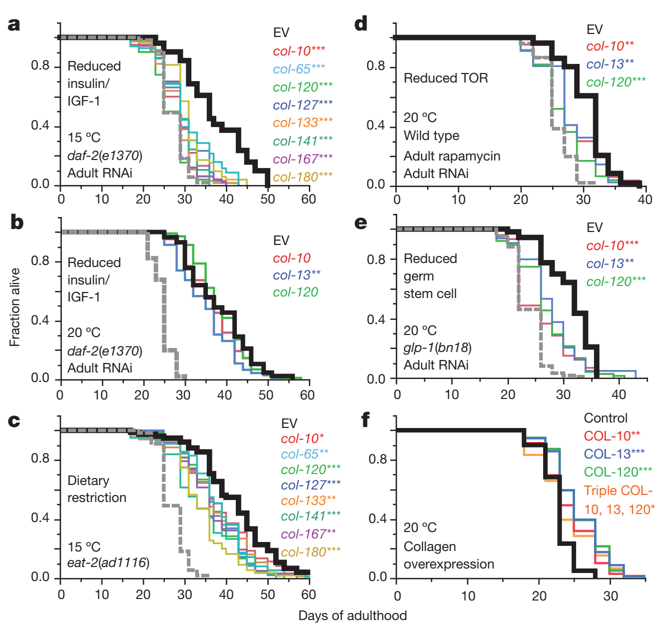


Figure 3 | Adulthood collagen expression promotes longevity. **a**, **b**, SKN-1-upregulated collagens are needed for *daf-2*(*e1370*) longevity at 15 but not 20 °C. **c**, Adulthood collagen knockdown reduced *eat-2*(*ad1116*) lifespan at 15 °C. Trial run in parallel with **a** and Extended Data Fig. 4c. **d**, Adulthood collagen expression is required for rapamycin lifespan extension. Rapamycin treatment and RNAi were initiated at adulthood day 1. **e**, Longevity from reduced germline stem cell number requires adult collagen expression. *glp-1*(*bn18*) was exposed to RNAi or empty RNAi vector control after downshift from the non-permissive temperature to 20 °C. In **a**–**e**, the grey dashed line shows the wild-type or control lifespan. **f**, Overexpression of collagens COL-10, COL-13, and COL-120 individually or in combination increased lifespan. $*P < 0.05$, $**P < 0.001$, $***P < 0.0001$ by log-rank. Statistics and additional lifespan data are in Extended Data Table 3 and Supplementary Table 13.

program may be controlled mainly by IIS acting outside the nervous system. The requirement for SKN-1 for lifespan extension is relieved under conditions that activate vestiges of the dauer developmental pathway in adults.

Analyses of how rIIS affects ageing have typically involved conditions that predispose to mild or even severe dauer-related traits (Supplementary Discussion), and would therefore allow *skn-1*-independent lifespan extension. We investigated the basis for dauer-independent rIIS longevity by identifying genes that are regulated by SKN-1 in *daf-2* mutants at 15 °C. At a false discovery rate of less than 3%, microarrays identified 429 genes with higher expression in *daf-2(−)* than *daf-2(−); skn-1(−)* animals (SKN-1-upregulated *daf-2(−)* genes), and 477 SKN-1-downregulated *daf-2(−)* genes, including direct and indirect SKN-1 targets (Extended Data Fig. 2a–e and Supplementary Table 3). Many of these genes affected lifespan as would be predicted by these expression patterns (Extended Data Fig. 2f–h and Supplementary Tables 4 and 5). Overlap with a dauer-expressed gene set was insignificant, as was overlap between SKN-1- and DAF-16-downregulated *daf-2(−)* genes (Extended Data Fig. 2i–k). However, many SKN-1-upregulated *daf-2(−)* genes were activated by DAF-16 (Extended Data Fig. 2j, l–t), which is also required for *daf-2* lifespan extension at 15 °C (ref. 17), indicating that SKN-1 responds to rIIS by functioning both in parallel to and independently of DAF-16.

SKN-1 has conserved functions in stress defence, protein homeostasis, and metabolism[12,18,19] and was required for *daf-2* oxidative stress resistance (Supplementary Table 6)[13], but only 40 out of 429 SKN-1-upregulated *daf-2(−)* genes had been identified under normal or stress conditions (Extended Data Fig. 3a–g and Supplementary Table 7)[18]. Unexpectedly, by far the most overrepresented functional group within the SKN-1-upregulated *daf-2(−)* gene set consisted of collagen genes, which seemed to be regulated by SKN-1 indirectly (Fig. 2a and Supplementary Tables 3, 8 and 9). In humans, collagens constitute about one-third of all protein and accumulate damage during ageing, leading to functional decline in tissues throughout the body[6,7]. *C. elegans* collagens form basement membranes as well as the cuticle, a complex structure that covers the animal, lines the buccal cavity, pharynx, and rectum, and becomes thickened and wrinkled with age[20]. The SKN-1-upregulated *daf-2(−)* collagens are of the type that forms the cuticle, but are expressed in multiple tissues (Extended Data Fig. 3h and Supplementary Table 9). Collagen production decreases in human skin during ageing[21], and 27 SKN-1-upregulated *daf-2(−)* collagens are among a set of genes that decline in expression as *C. elegans* ages[22] (Supplementary Table 10). These and other collagens were prominently upregulated in each of 20 *C. elegans* longevity-associated gene sets we examined (Extended Data Table 2 and Supplementary Table 10). Moreover, in mice extracellular matrix (ECM) genes were overrepresented in some longevity or Nrf2-dependent sets (Supplementary Tables 11 and 12), and *in silico* analysis of longevity-associated genes identified a predicted ECM network[23]. The possible significance of these expression signatures has not been explored.

We investigated the functional importance of specific SKN-1-upregulated *daf-2(−)* collagen genes that decline during ageing, and are upregulated in other longevity-associated gene sets (Extended Data Table 2). SKN-1 increased expression of these genes during adulthood, and delayed their age-related decline in expression in response to multiple interventions that promote longevity: *daf-2* RNAi, rapamycin (mTOR kinase inhibitor[24]), the dietary restriction model *eat-2*, and inhibition of germ cell proliferation (*glp-1(−)*)[1] (Fig. 2b–g and Extended Data

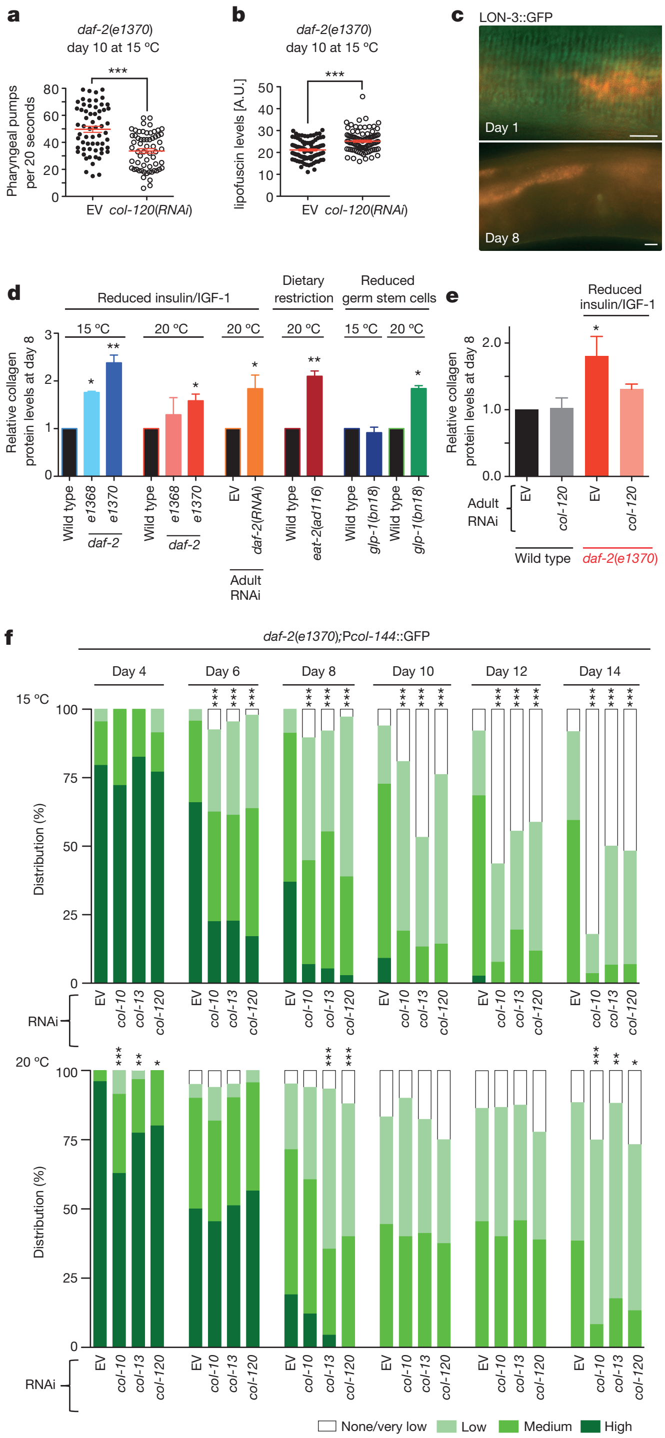


Figure 4 | Importance of ECM remodelling for longevity assurance. **a**, **b**, Adulthood collagen expression is required for rIIS to delay appearance of ageing markers. The same animals were scored in each panel ($N > 60$). Each dot represents an animal; two merged trials; $^{***}P < 0.0001$ determined with unpaired *t*-test, two-tailed. **c**, Disappearance of the LON-3 collagen from the cuticle during ageing. Typical animals at the indicated days of adulthood are shown. Midsections from representative LON-3::GFP (green fluorescent protein) adults are shown, ventral side down, anterior to the left; scale bar, 10 μm. **d**, Interventions that increase longevity induce adulthood ECM deposition. Total collagen in day 8 adults is indicated by hydroxyproline content. Initiation of *daf-2* RNAi was at day 1. The *glp-1(bn18)* mutants were kept at the permissive temperature (15 °C) or shifted to 25 °C until day 1 of adulthood then kept at 20 °C. **e**, Loss of a single collagen interferes with rIIS-induced collagen deposition. In **d**, **e**, $N > 3{,}000$ per sample. Data are mean ± s.e.m. $^{*}P < 0.05$ relative to control, by one sample *t*-test, two-tailed, hypothetical mean of 1. **f**, Dependence of a collagen promoter (*col-144*) on adulthood expression of other SKN-1-upregulated collagens in *daf-2(e1370)* under dauer-independent conditions. Scoring is described in Extended Data Fig. 7d. RNAi initiated at day 1 of adulthood had a much more severe effect at 15 °C (upper panel) than 20 °C (lower panel), starting at day 6. $N > 60$ for each condition, one representative trial is shown, with P value by χ^2 test ($^{*}P < 0.05$, $^{**}P < 0.001$, $^{***}P < 0.0001$).

Figs 3i–k and 4a, b). Adulthood knockdown of these collagen genes did not affect wild-type lifespan, but dramatically reduced longevity of the canonical *daf-2* class 2 mutant *e1370* at 15 °C but not 20 °C (Fig. 3a, b, Extended Data Fig. 4c, Extended Data Table 3 and Supplementary Table 13), at which *skn-1* is dispensable for longevity (see above). Additionally, knockdown of these collagens significantly reduced lifespan extension from *daf-2* RNAi at 20 °C, and from other *skn-1*-dependent[14,24,25] longevity interventions (Fig. 3c–e, Extended Data Fig. 4d, Table 3 and Supplementary Table 13). Most of these genes include regions related to other collagens, but *col-120* is unique (Supplementary Table 14), and at 15 °C *daf-2*(*e1370*) but not wild-type lifespan was reduced by the collagen mutation *dpy-1*(*e1*) (Extended Data Fig. 4e and Supplementary Table 13). Lack of a single critical collagen can therefore impair lifespan extension. At 15 °C, *daf-2*(*e1370*) lifespan was also decreased by adulthood knockdown of certain extracellular protease genes from the SKN-1-upregulated *daf-2*(−) set, or other genes important for cuticle formation (Extended Data Fig. 4f and Supplementary Tables 13 and 15). Remarkably, transgenic overexpression of key collagens from the SKN-1-upregulated *daf-2*(−) gene set but not other collagens modestly but consistently increased lifespan (Fig. 3f and Supplementary Table 13). Adulthood SKN-1-dependent expression of particular collagen and ECM genes therefore promotes lifespan extension in diverse pathways that slow *C. elegans* ageing.

Adulthood collagen RNAi did not affect body size, detectably impair cuticle function, or increase markers of various stresses (Extended Data Figs 5a–v and 6a–i). Collagen RNAi sensitized to exogenous oxidative stress, however, and increased the prominence of ageing markers in *daf-2* mutants at 15 °C, and in rapamycin-treated animals (Fig. 4a, b, Extended Data Fig. 6j–m and Supplementary Table 16). Apparently, knockdown of these collagens interfered with the capacity of these interventions to delay ageing.

ECM gene upregulation might allow ECM remodelling to occur in adults. During ageing the collagens LON-3 and ROL-6 decline in expression[22] and largely disappear from the cuticle (Fig. 4c and Extended Data Figs 4a and 7a), indicating that *C. elegans* ECM proteins turn over. Adulthood *daf-2* RNAi and other anti-ageing interventions increased total collagen in older *C. elegans* (Fig. 4d), indicating deposition of new ECM. This also occurred in *daf-2*(*e1370*) (class 2) at 20 °C, even though by adulthood day 8 expression of SKN-1 upregulated *daf-2*(−) collagens was not generally maintained in older *daf-2*(*e1370*) adults under these conditions (Fig. 4d and Extended Data Fig. 7b, c). Perhaps different genes might promote ECM remodelling under dauer-predisposed conditions, consistent with dauers having a distinct cuticle structure (Supplementary Discussion).

Longevity interventions delay ageing by acting through non-cell-autonomous signalling pathways[1]. Adulthood *col-120* knockdown reduced total *daf-2* collagen levels (Fig. 4e), implying that individual collagens and the ECM influence these pathways. Adulthood collagen RNAi also inhibited SKN-1-responsive gene expression in adults that would otherwise be long-lived (Fig. 4f and Extended Data Fig. 7d–g), possibly explaining the importance of these collagens for oxidative stress resistance. These longevity interventions therefore require adulthood expression of particular ECM genes to maintain their beneficial regulatory program. Why would diverse longevity interventions induce and depend upon ECM remodelling? Under conditions of low nutrient availability, it might be advantageous to allocate resources towards ECM maintenance. The ECM also may directly affect signalling that orchestrates these longevity pathways, consistent with studies in other systems that identified signalling functions of collagens, and critical effects of the ECM on signalling pathways[26–28].

We determined that in adult animals rIIS can activate a longevity program that is distinguished from the dauer developmental pathway by its lack of dauer-like traits, and its dependence upon *skn-1* and SKN-1-dependent collagens (Fig. 1f). Further analyses will determine which rIIS longevity mechanisms are linked to the dauer program, and which are dauer-independent and possibly more broadly involved in pathways that promote longevity. Considerable effort has been devoted to enhancing collagen function to maintain youthful human skin during ageing[29]. By demonstrating that increased collagen expression is a shared feature of multiple conserved longevity pathways, our results suggest strategies for promoting ECM function that may be widely applicable. The long-lived naked mole rat is remarkably cancer resistant, at least in part because it produces a uniquely dense hyaluronan, an ECM component[30]. Our results suggest that functional enhancement of the ECM may be generally important for longevity assurance per se. We speculate that interventions that promote collagen and ECM function systemically are likely to be beneficial in human chronic disease and ageing.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 October 2013; accepted 27 October 2014.
Published online 15 December 2014.

1. Kenyon, C. J. The genetics of ageing. *Nature* **464,** 504–512 (2010).
2. Shore, D. E. & Ruvkun, G. A cytoprotective perspective on longevity regulation. *Trends Cell Biol.* **23,** 409–420 (2013).
3. Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153,** 1194–1217 (2013).
4. Flurkey, K., Papaconstantinou, J., Miller, R. A. & Harrison, D. E. Lifespan extension and delayed immune and collagen aging in mutant mice with defects in growth hormone production. *Proc. Natl Acad. Sci. USA* **98,** 6736–6741 (2001).
5. Wilkinson, J. E. *et al.* Rapamycin slows aging in mice. *Aging Cell* **11,** 675–682 (2012).
6. Myllyharju, J. & Kivirikko, K. I. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet.* **20,** 33–43 (2004).
7. Toyama, B. H. & Hetzer, M. W. Protein homeostasis: live long, won't prosper. *Nature Rev. Mol. Cell Biol.* **14,** 55–61 (2013).
8. Partridge, L. & Harvey, P. H. Gerontology. Methuselah among nematodes. *Nature* **366,** 404–405 (1993).
9. McElwee, J. J., Schuster, E., Blanc, E., Thomas, J. H. & Gems, D. Shared transcriptional signature in *Caenorhabditis elegans* dauer larvae and long-lived *daf-2* mutants implicates detoxification system in longevity assurance. *J. Biol. Chem.* **279,** 44533–44543 (2004).
10. Gems, D. *et al.* Two pleiotropic classes of *daf-2* mutation affect larval arrest, adult behavior, reproduction and longevity in *Caenorhabditis elegans*. *Genetics* **150,** 129–155 (1998).
11. Arantes-Oliveira, N., Berman, J. R. & Kenyon, C. Healthy animals with extreme longevity. *Science* **302,** 611 (2003).
12. Sykiotis, G. P. & Bohmann, D. Stress-activated cap'n'collar transcription factors in aging and human disease. *Sci. Signal* **3,** re3 (2010).
13. Tullet, J. M. *et al.* Direct inhibition of the longevity-promoting factor SKN-1 by insulin-like signaling in *C. elegans*. *Cell* **132,** 1025–1038 (2008).
14. Robida-Stubbs, S. *et al.* TOR signaling and rapamycin influence longevity by regulating SKN-1/Nrf and DAF-16/FoxO. *Cell Metab.* **15,** 713–724 (2012).
15. Libina, N., Berman, J. R. & Kenyon, C. Tissue-specific activities of *C. elegans* DAF-16 in the regulation of lifespan. *Cell* **115,** 489–502 (2003).
16. Wolkow, C. A., Kimura, K. D., Lee, M. S. & Ruvkun, G. Regulation of *C. elegans* life-span by insulinlike signaling in the nervous system. *Science* **290,** 147–150 (2000).
17. Narasimhan, S. D. *et al.* PDP-1 links the TGF-β and IIS pathways to regulate longevity, development, and metabolism. *PLoS Genet.* **7,** e1001377 (2011).
18. Oliveira, R. P., Porter Abate, J. & Dilks, K. *et al.* Condition-adapted stress and longevity gene regulation by *Caenorhabditis elegans* SKN-1/Nrf. *Aging Cell* **8,** 524–541 (2009).
19. Pang, S., Lynn, D. A., Lo, J. Y., Paek, J. & Curran, S. P. SKN-1 and Nrf2 couples proline catabolism with lipid metabolism during nutrient deprivation. *Nature Commun.* **5,** 5048 (2014).
20. Herndon, L. A. *et al.* Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*. *Nature* **419,** 808–814 (2002).
21. Varani, J. *et al.* Decreased collagen production in chronologically aged skin: roles of age-dependent alteration in fibroblast function and defective mechanical stimulation. *Am. J. Pathol.* **168,** 1861–1868 (2006).
22. Budovskaya, Y. V. *et al.* An elt-3/elt-5/elt-6 GATA transcription circuit guides aging in *C. elegans*. *Cell* **134,** 291–303 (2008).
23. Argmann, C. *et al.* Pparγ2 is a key driver of longevity in the mouse. *PLoS Genet.* **5,** e1000752 (2009).
24. Johnson, S. C., Rabinovitch, P. S. & Kaeberlein, M. mTOR is a key modulator of ageing and age-related disease. *Nature* **493,** 338–345 (2013).
25. Vilchez, D. *et al.* RPN-6 determines *C. elegans* longevity under proteotoxic stress conditions. *Nature* **489,** 263–268 (2012).
26. Seeger-Nukpezah, T. & Golemis, E. A. The extracellular matrix and ciliary signaling. *Curr. Opin. Cell Biol.* **24,** 652–661 (2012).
27. Munger, J. S. & Sheppard, D. Cross talk among TGF-β signaling pathways, integrins, and the extracellular matrix. *Cold Spring Harb. Perspect. Biol.* **3,** a005017 (2011).
28. Fu, H. L. *et al.* Discoidin domain receptors: unique receptor tyrosine kinases in collagen-mediated signaling. *J. Biol. Chem.* **288,** 7430–7437 (2013).

29. Baumann, L. Skin ageing and its treatment. *J. Pathol.* **211,** 241–251 (2007).
30. Tian, X. *et al.* High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature* **499,** 346–349 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Kenyon, S. Mitani, and J. Shim for strains, P. Sengupta for dauer pheromone, C. Obieglo, L. Moronetti, M. Bland, and K. Patel for assistance, and J. Apfeld, E. Greer, C. Kenyon, W. Mair, and Blackwell laboratory members for discussions or comments on the manuscript. Some strains were provided by the Caenorhabditis Genetics Center, which is funded by the National Institutes of Health Office of Research Infrastructure Programs (P40 OD010440). The work was supported by funding from the National Institutes of Health to T.K.B. (GM062891), C.T.M. (New Innovator), and J.P.A. (5T32DK007260), a Diabetes Research Center award to the Joslin Diabetes Center (P30DK036836), and fellowships from the National Science Foundation to J.N.L., and the Swiss National Science Foundation (PBSKP3_140135) to C.Y.E.

Author Contributions All authors participated in designing the experiments, and analysing and interpreting the data. J.N.L. and J.P.A. obtained samples for microarray analysis, performed the microarray experiments, analysed the expression profiling data, and performed the lifespan studies in Extended Data Fig. 2f–h and Supplementary Table 4. C.Y.E. performed all other experiments. C.Y.E. and T.K.B. wrote the manuscript in consultation with the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.K.B. (keith.blackwell@joslin.harvard.edu) or C.T.M. (ctmurphy@princeton.edu).

METHODS

Strains. *C. elegans* strains were maintained on NGM plates and OP50 *Escherichia coli* bacteria at 20 °C as described[31], except that *daf-2* mutants (and corresponding controls for a given assay) were maintained at 15 °C unless otherwise noted. The wild-type strain was N2 Bristol[31]. Mutant strains used are described in Wormbase (www.wormbase.org). LGI: *daf-16*(*mgDf47*, *mu86*); LGII: *eat-2*(*ad1116*); LGIII: *daf-2*(*e1368*, *e1370*, and *m596*), *rrf-3*(*pk1462*), *glp-1*(*bn18*); and LGIV: *eri-1*(*mg366*), *skn-1*(*tm3411*, *zu67*, *zu129*, and *zu135*). LGX: *lin-15B*(*n744*). The following transgenic lines were used: *jgIs5* [ROL- 6::GFP;TTX-3::GFP][32], BC12533 *dpy-5*(*e907*); *sEx12533* [P*col-89*::GFP; *dpy-5*(+)][33], CF1660 *daf-16*(*mu86*); *daf-2*(*e1370*); *muIs84* [P*sod-3*::GFP; pRF4 *rol-6*(*su1006gf*)]; *muEx211* [P*ges-1*::DAF-16::GFP; pRF4 *rol-6*(*su1006gf*)][15], CL2166 *dvIs19* [P*gst-4*::GFP; pRF4 *rol-6*(*su1006gf*)][34], EE86 *mup-4*(*mg36*); *upIs1* [MUP-4::GFP; pRF4 *rol-6*(*su1006gf*)][35], HT1883 *daf-16*(*mgDf50*); *daf-2*(*e1370*) *unc-119*(*ed3*); *lpIs14* [P*daf-16*::DAF-16f::GFP + *unc-119*(+)][36], IG274 *frIs7* [P*col-12*::DsRed; P*nlp-29*::GFP][37], LD001 *ldIs007* [P*skn-1*::SKN-1b/c::GFP; pRF4 *rol-6*(*su1006gf*)][38], MH2051 *kuIs55* [LON-3::GFP; *unc-119*(+)][39], SJ4005 *zcIs4* [Phsp-4::GFP; *lin-15*(+)][40], SJ4103 *zcIs14* [myo-3::GFP(mit)][41], TB1682 *chEx1682* [QUA-1::GFP; pRF4 *rol-6*(*su1006gf*)][42], TJ356 *zIs356* [P*daf-16*::DAF-16a/b::GFP; pRF4 *rol-6*(*su1006gf*)][43], TP12 *kaIs12* [COL-19::GFP][44].

Construction of transgenic lines. To construct the collagen overexpression transgenes, the genomic region of each gene, including approximately 3 kilobases (kb) of promoter, the coding region, and 3′ untranslated region sequences that encompass at least two predicted cleavage/polyadenylation sites, were amplified by PCR. These PCR products were injected at 50 ng μl^{-1} together with 100 ng μl^{-1} of pRF4 *rol-6*(*su1006gf*) into wild-type (N2) animals. For the triple collagen gene transgenic line (*ldEx111*), 50 ng μl^{-1} each of PCR products for *col-10*, *col-13*, *col-120* were injected together with 50 ng μl^{-1} of pRF4 *rol-6*(*su1006gf*). For the control line (*ldEx102*), pBluescript KS(+) 50 ng μl^{-1} was injected along with 100 ng μl^{-1} of pRF4 *rol-6*(*su1006gf*). Lines were isolated from at least two independent transgenic P0 animals. For *col-10*, a 4.4 kb genomic region was amplified using the primers 5′-CC ACCAACAACTCCATCCACC-3′ and 5′-GTAAAGTGGGCAGGCCGTAG-3′. The resulting transgenic lines were *ldEx103* and *ldEx104*. For *col-13*, a 4.3 kb genomic region was amplified using the primers 5′-TAGCCCAAGTCTGACCGA AG-3′ and 5′-CGGATCTTCCCAACCAGGAG-3′. The resulting transgenic lines were *ldEx105*, *ldEx106*, *ldEx107*, and *ldEx108*. For *col-120*, a 4.4 kb genomic region was amplified using the primers 5′-CAATATGACCCGAGGCGCTG-3′ and 5′-CGCCAGAATCGTAAGGCTCC-3′. The resulting transgenic lines were: *ldEx109* and *ldEx110*. Transgene overexpression levels were determined by qPCR of 1-day-old adults.

Scoring of phenotypic experiments. No statistical methods were used in choosing sample sizes. In analyses of fluorescent reporters, either all or representative trials were scored blindly. All other phenotypic assays were not scored blindly.

Body length measurements. Animals were maintained at 15 °C and either kept at 15 °C, or shifted to 25 °C at the first day of adulthood. At day 3 of adulthood, animals were mounted on 2% agar pads, immobilized with 0.06% tetramisole and images were taken at ×10 magnification with a Zeiss Axioskop 2 microscope and a Zeiss AxioCam HRc digital camera. Body lengths were measured by placing a line through the middle of the body starting from head to tail using Zeiss AxioVision version 4.8.2.0 (Extended Data Fig. 1d).

Lifespan assays. Strains were age-synchronized by picking larval stage 4 (L4) animals onto fresh OP50 plates, then day 1 adults were placed on either OP50 or RNAi plates containing 50 μM 5-fluoro-2′deoxyuridine (FUdR), unless otherwise indicated, and assayed either at 15, 20, or 25 °C as described in ref. 14. All lifespans were plotted with L4 as time-point = 0. For *glp-1*(*bn18*) lifespans, wild type (N2) and *glp-1*(*bn18*) were maintained at 15 °C, then shifted to 25 °C at the mid-L1 stage as described in ref. 45. At the first day of adulthood they were placed on plates containing FUdR and RNAi bacteria for lifespan assay at 20 °C (Fig. 4g and Extended Data Table 3). For rapamycin lifespans, 1-day-old animals were placed on plates containing FUdR, RNAi bacteria, and either rapamycin (100 μM) dissolved in 0.2% dimethylsulphoxide (DMSO) or 0.2% DMSO control as described in ref. 14. Lifespan was determined at 20 °C (Fig. 3d, Extended Data Table 3 and Supplementary Table 13). For dauer pheromone experiments, day 1 adults were placed on plates containing FUdR, RNAi bacteria, and either crude dauer pheromone (a gift from P. Sengupta) dissolved in 6% ethanol, or 6% ethanol control as described in ref. 46. Those lifespans were determined at 25 °C (Fig. 1e, Extended Data Table 1 and Supplementary Table 2). Animals were classified as dead if they failed to respond to prodding. Exploded or bagged animals were excluded from the statistics. The estimates of survival functions were calculated using the product-limit (Kaplan–Meier) method. The log-rank (Mantel–Cox) method was used to test the null hypothesis and calculate *P* values (JMP software version 9.0.2.).

Scoring of transgenic protein nuclear accumulation or expression. Nuclear accumulation of SKN-1 that was expressed from the *SKN-1bc::GFP* transgene (LD001 strain), which encodes two of the three SKN-1 isoforms, was scored blindly after mounting on slides essentially as in ref. 14 (Extended Data Fig. 1k–n). Scoring was as follows: none, no GFP observed in nuclei; low, some nuclei showed GFP; medium, more than half of the nuclei showed GFP; high, all intestinal nuclei showed GFP. Nuclear accumulation of DAF-16a/b::GFP (*zIs356*) was scored as described in ref. 47 (Extended Data Fig. 1n). Nuclear accumulation of DAF-16f::GFP (*lpIs14*) was scored as follows: none, no GFP observed in nuclei; medium, more than half of the nuclei showed GFP; high, all intestinal nuclei showed GFP (Extended Data Fig. 1o). For P*col-12*::dsRED, P*col-144*::GFP, P*gst-4*::GFP, and P*hsp-4*::GFP, one-day adult animals were placed on RNAi and 3 and/or 7 days later the green or red fluorescence intensity was scored by using a Zeiss AxioSKOP2 microscope. Green or red fluorescence was categorized in none/very low, low, medium, or high intensity and was scored blindly (Fig. 4f and Extended Data Figs 3j–k, 6h–j and 7d–g).

RNAi. RNAi clones were picked from the Ahringer[48] or Vidal[49] libraries. Cultures were grown overnight in lysogeny broth with 12.5 μg ml^{-1} tetracycline and 100 μg ml^{-1} ampicillin, diluted to an attenuance ($D_{600\,nm}$) of 1, and induced with 1 mM IPTG. This culture was seeded onto NGM agar plates containing tetracycline, ampicillin, and additional IPTG. Empty vector plasmid pL4440 was used as control. For double RNAi, clones were grown separately in parallel and after spin-down equal amounts of two clones were mixed and spread on plates.

RNA isolation for microarray analysis. After a timed egg-lay on HT115 *E. coli*, *daf-2*(*e1368*) and *daf-2*(*e1368*); *skn-1*(*zu67*) or *daf-2*(*e1370*) and *daf-2*(*e1370*); *skn-1*(*zu67*) worms were grown at 15 °C until the late L4 stage. Approximately 200 worms were collected and washed three times in M9 buffer[31] to remove bacteria. TriReagent (Sigma) was added, and samples were snap frozen in liquid nitrogen. Total RNA was isolated using TriReagent and an RNA purification column (RNA-easy, Qiagen). RNA quality was determined by visualization of 28S and 18S ribosomal RNA bands on a denaturing formaldehyde gel, or an RNase-free 1.5–2% agarose TBE gel.

RNA preparation, hybridization and data collection for microarray experiments. RNA (325 ng) was linearly amplified and labelled using the Agilent Low RNA Input Linear Amplification Kit, with Cy3- or Cy5-CTP (Perkin Elmer), then RNA copies were hybridized on Agilent 4×44k *C. elegans* arrays. A dye swap replicate was performed for each set of biological replicate samples as previously described[18]. Data were extracted with Agilent Feature Extraction software and submitted to the Princeton University Microarray Database (PUMAdb) for storage and filtering (https://puma.princeton.edu). These microarray data are publicly available at PUMAdb.

Microarray analysis. Data were filtered to remove spots that were not above background intensity in both channels, and replicate spots within each array were averaged. Genes for which more than 20% of data were missing across replicates were removed from further analysis. One-class SAM analysis was used to identify genes that were significantly up- or downregulated across all replicates in a set[50]. Expression profiles were clustered using Cluster 3.0 (ref. 51) and visualized using Java TreeView[52]. Up- and downregulated genes identified by SAM analysis were submitted to DAVID[53] to identify overrepresented functional annotations. Annotations used were Gene Ontology (GO) Biological Process FAT (GO BP, filtered by DAVID to remove the broadest GO terms), GO Molecular Function, Kegg Pathway, and Interpro Protein Domains. The Benjamini test for multiple hypothesis testing was applied to *P* values. Up- and downregulated genes were also submitted to GOToolBox to perform a hypergeometric test using the Benjamini–Hochberg correction. Enriched GO terms were submitted to ReviGO to remove redundant terms. Co-occurrence between our data sets and previously published data sets was visualized with GeneVenn[54] and BioInfoRx Area-Proportional Venn Diagram.

Motif analysis. We used two distinct algorithms, Weeder[55] and FIRE[56], to perform an unbiased search for overrepresented sequences in the promoters of SKN-1-regulated genes that were identified by SAM. We submitted upstream sequences (1,000 base pairs (bp)) to Weeder and performed a scan for motifs of length 6 and 8 ('normal' scan mode). FIRE was run using default parameters, with all genes partitioned into three groups to identify motifs that were informative about each group: SKN-1-upregulated, SKN-1-downregulated, and background. To search in a directed manner for occurrence of the consensus SKN-1 binding motif, we used RSATools[57] to search the 600 bp upstream of up- and downregulated targets for the SKN-1 binding motif (WWTRTCAT). For comparison with the percentage of promoters in a random sample of genes that would be expected to contain the SKN-1 motif, we searched for the motif in 10,000 random samplings of gene promoter sets of equal size to the number of up- or downregulated genes, to determine a distribution empirically. To calculate a *P* value, we *z*-transformed the percentage of SKN-1 target promoters ($z = (\%_{SKN\text{-}1} - \mu)/\sigma$), where μ and σ are the mean and standard deviation of the distribution.

qPCR assays. For validation of the microarray data with *skn-1* and *daf-16* mutants, *C. elegans* were allowed to lay eggs for 3–4 h on RNAi plates. After 2–4 days (depending upon the temperature and strain), 200 L4 worms were harvested (15 °C for Extended Data Fig. 2c; 20 °C Extended Data Fig. 2m–t). For adult RNAi, 1-day-old adults were placed on RNAi plates and 3–8 days later 100–200 worms were

harvested (Fig. 2a–c and Extended Data Fig. 4b). For rapamycin treatment, 1-day-old animals were placed on plates containing rapamycin (100 μM) dissolved in 0.2% DMSO or in 0.2% DMSO control as described in ref. 14 and 3 days later mRNA was harvested for qPCR (Extended Data Fig. 3l–m). For the *glp-1* experiment, *glp-1*(*bn18*) or wild-type (N2) animals were maintained at 15 °C and L2 worms were upshifted to 25 °C. Day 1 adults were placed on L4440 (empty vector RNAi) plates at 20 °C and 3 days later 200 worms were harvested (Extended Data Fig. 3p–q). RNA was isolated with Trizol (TRI REAGENT Sigma), DNase-treated, and cleaned over a column (RNA Clean & Concentrator^{TM} ZYMO Research). First-strand complementary DNA (cDNA) was synthesized in duplicate from each sample (Invitrogen SuperScript III). SYBR green was used to perform qPCR (ABI 7900). For each primer set, a standard curve from genomic DNA accompanied the duplicate cDNA samples[58]. mRNA levels relative to N2 control were determined by normalizing to the number of worms and the geometric mean of three reference genes (*cdc-42*, *pmp-3*, and Y45F10D.4 (ref. 59)). Primer sequences are listed in Supplementary Table 17. Except for *col-12/13*, primers bound uniquely to the corresponding gene transcript (Supplementary Tables 14 and 17). At least two biological replicates were examined for each sample. For statistical analysis, one sample *t*-test, two-tailed, hypothetical mean of 1, was used for comparison using Prism 4.0a software (GraphPad).

Oxidative stress assays. In oxidative stress assays, day 1 *daf-2* or *skn-1* adults were placed in 5 mM sodium arsenite (in 1 ml H_2O) at 20 °C and scored for survival hourly (Supplementary Table 7). For RNAi oxidative stress assays, wild-type (N2) or *daf-2*(*e1370*) day 1 adults were placed on RNAi plates at 15 °C, and 3 days later animals were placed either on plates containing 15.4 mM t-BOOH and scored hourly at 20 °C, or in 5 mM sodium arsenite (in 1 ml M9 buffer) and scored after 21 h (N2) or 30 h (*daf-2*) at 20 °C (Extended Data Fig. 6j–l and Supplementary Table 16).

Age-related phenotypic marker and body-size assays. Age-related phenotypes were described in ref. 60. One-day-old animals were placed on RNAi food until day 10 of adulthood and the following phenotypes were scored. (1) Pharyngeal pumping was determined by counting grinder movements in 20 s intervals when the animals were placed on food (Fig. 4a and Extended Data Fig. 6m). (2) Lipofuscin levels were determined by mounting animals onto slides and taking bright-field and 4′,6-diamidino-2-phenylindole (DAPI) channel pictures with a Zeiss Imager M2 microscope. Blue fluorescence from the DAPI channel pictures were analysed in Image J (http://imagej.nih.gov/ij/) by selecting the intestine and measuring the mean grey value minus the background (Fig. 4b). (3) The body size was determined from bright-field images by drawing a line through the middle of the worm from anterior to posterior by using Zeiss Zen 2012 software (Extended Data Fig. 6a).

Collagen assays. Synchronized L1 larvae were placed on 10 cm NGM plates containing OP50 bacteria at 15, 20, or 25 °C and monitored for development to the L4 stage. After an additional day, day 1 adults were either harvested for the assay (Extended Data Fig. 7b), or placed on either 10 cm OP50 or RNAi plates containing 50 μM FUdR and maintained at the corresponding temperature. At day 8 of adulthood, the remaining animals were harvested (Fig. 4d, e). In each case, the animals were washed three times with M9, the number of worms was determined, and at least 3,000 worms per strain and condition were used for the assay. Collagen levels were determined using the QuickZyme Biosciences Total Collagen Kit (QZBTOT COL1), which detects hydroxyproline[61], according to the manufacturer's instructions.

Barrier function assay. One-day-old adults were placed on RNAi food and at day 9 were harvested, washed three times with M9 and incubated in 1 μg ml^{-1} Hoechst (Hoechst 33342, which is cuticle-impermeable but membrane-permeable) for 15 min in darkness at room temperature (22 °C). The animals were then washed three times in M9, allowed to recover for 10 min on plates with food, and mounted for microscopy (Extended Data Fig. 6b, c; method adapted from ref. 62).

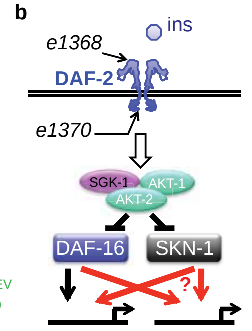
31. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77,** 71–94 (1974).
32. Kim, T. H. *et al.* Tyrosylprotein sulfotransferase regulates collagen secretion in *Caenorhabditis elegans*. *Mol. Cells* **29,** 413–418 (2010).
33. McKay, S. J. *et al.* Gene expression profiling of cells, tissues, and developmental stages of the nematode *C. elegans*. *Cold Spring Harb. Symp. Quant. Biol.* **68,** 159–169 (2003).
34. Link, C. D. & Johnson, C. J. Reporter transgenes for study of oxidant stress in *Caenorhabditis elegans*. *Methods Enzymol.* **353,** 497–505 (2002).
35. Hong, L. *et al.* MUP-4 is a novel transmembrane protein with functions in epithelial cell adhesion in *Caenorhabditis elegans*. *J. Cell Biol.* **154,** 403–414 (2001).
36. Kwon, E. S., Narasimhan, S. D., Yen, K. & Tissenbaum, H. A. A new DAF-16 isoform regulates longevity. *Nature* **466,** 498–502 (2010).
37. Pujol, N. *et al.* Distinct innate immune responses to infection and wounding in the *C. elegans* epidermis. *Curr. Biol.* **18,** 481–489 (2008).
38. An, J. H. & Blackwell, T. K. SKN-1 links *C. elegans* mesendodermal specification to a conserved oxidative stress response. *Genes Dev.* **17,** 1882–1893 (2003).
39. Suzuki, Y., Morris, G. A., Han, M. & Wood, W. B. A cuticle collagen encoded by the *lon-3* gene may be a target of TGF-β signaling in determining *Caenorhabditis elegans* body shape. *Genetics* **162,** 1631–1639 (2002).
40. Calfon, M. *et al.* IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. *Nature* **415,** 92–96 (2002).
41. Benedetti, C., Haynes, C. M., Yang, Y., Harding, H. P. & Ron, D. Ubiquitin-like protein 5 positively regulates chaperone gene expression in the mitochondrial unfolded protein response. *Genetics* **174,** 229–239 (2006).
42. Hao, L. *et al.* The hedgehog-related gene *qua-1* is required for molting in *Caenorhabditis elegans*. *Dev. Dyn.* **235,** 1469–1481 (2006).
43. Henderson, S. T. & Johnson, T. E. *daf-16* integrates developmental and environmental inputs to mediate aging in the nematode *Caenorhabditis elegans*. *Curr. Biol.* **11,** 1975–1980 (2001).
44. Thein, M. C. *et al. Caenorhabditis elegans* exoskeleton collagen COL-19: an adult-specific marker for collagen modification and assembly, and the analysis of organismal morphology. *Dev. Dyn.* **226,** 523–539 (2003).
45. Arantes-Oliveira, N., Apfeld, J., Dillin, A. & Kenyon, C. Regulation of life-span by germ-line stem cells in *Caenorhabditis elegans*. *Science* **295,** 502–505 (2002).
46. Neal, S. J., Kim, K. & Sengupta, P. Quantitative assessment of pheromone-induced dauer formation in *Caenorhabditis elegans*. *Methods Mol. Biol.* **1068,** 273–283 (2013).
47. Curran, S. P. & Ruvkun, G. Lifespan regulation by evolutionarily conserved genes essential for viability. *PLoS Genet.* **3,** e56 (2007).
48. Kamath, R. S., Martinez-Campos, M., Zipperlen, P., Fraser, A. G. & Ahringer, J. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol.* **2,** research0002 (2001).
49. Rual, J. F. *et al.* Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome Res.* **14,** 2162–2168 (2004).
50. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98,** 5116–5121 (2001).
51. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95,** 14863–14868 (1998).
52. Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20,** 3246–3248 (2004).
53. Dennis, G. Jr *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4,** 3 (2003).
54. Pirooznia, M., Nagarajan, V. & Deng, Y. GeneVenn - a web application for comparing gene lists using Venn diagrams. *Bioinformation* **1,** 420–422 (2007).
55. Pavesi, G., Mereghetti, P., Mauri, G. & Pesole, G. Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res.* **32,** W199–W203 (2004).
56. Elemento, O., Slonim, N. & Tavazoie, S. A universal framework for regulatory element discovery across all genomes and data types. *Mol. Cell* **28,** 337–350 (2007).
57. Thomas-Chollier, M. *et al.* RSAT: regulatory sequence analysis tools. *Nucleic Acids Res.* **36,** W119–W127 (2008).
58. Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D. L. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature Struct. Mol. Biol.* **15,** 71–78 (2008).
59. Hoogewijs, D., Houthoofd, K., Matthijssens, F., Vandesompele, J. & Vanfleteren, J. R. Selection and validation of a set of reliable reference genes for quantitative sod gene expression analysis in *C. elegans*. *BMC Mol. Biol.* **9,** 9 (2008).
60. Huang, C., Xiong, C. & Kornfeld, K. Measurements of age-related changes of physiological processes that predict lifespan of *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **101,** 8084–8089 (2004).
61. Zito, E., Hansen, H. G., Yeo, G. S., Fujii, J. & Ron, D. Endoplasmic reticulum thiol oxidase deficiency leads to ascorbic acid depletion and noncanonical scurvy in mice. *Mol. Cell* **48,** 39–51 (2012).
62. Kage-Nakadai, E. *et al.* Two very long chain fatty acid acyl-CoA synthetase genes, *acs-20* and *acs-22*, have roles in the cuticle surface barrier in *Caenorhabditis elegans*. *PLoS ONE* **5,** e8857 (2010).
63. Hertweck, M., Gobel, C. & Baumeister, R. C. *elegans* SGK-1 is the critical component in the Akt/PKB kinase complex to control stress response and life span. *Dev. Cell* **6,** 577–588 (2004).
64. Patel, D. S. *et al.* Clustering of genetically defined allele classes in the *Caenorhabditis elegans* DAF-2 insulin/IGF-1 receptor. *Genetics* **178,** 931–946 (2008).
65. Dillin, A., Crawford, D. K. & Kenyon, C. Timing requirements for insulin/IGF-1 signaling in *C. elegans*. *Science* **298,** 830–834 (2002).
66. Apfeld, J. & Kenyon, C. Cell nonautonomy of *C. elegans daf-2* function in the regulation of diapause and life span. *Cell* **95,** 199–210 (1998).
67. Timmons, L., Tabara, H., Mello, C. C. & Fire, A. Z. Inducible systemic RNA silencing in *Caenorhabditis elegans*. *Mol. Biol. Cell* **14,** 2972–2983 (2003).
68. Kennedy, S., Wang, D. & Ruvkun, G. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427,** 645–649 (2004).
69. Blackwell, T. K., Bowerman, B., Priess, J. R. & Weintraub, H. Formation of a monomeric DNA binding domain by Skn-1 bZIP and homeodomain elements. *Science* **266,** 621–628 (1994).
70. Niu, W. *et al.* Diverse transcription factor binding features revealed by genome-wide ChIP-seq in *C. elegans*. *Genome Res.* **21,** 245–254 (2011).
71. Simmer, F. *et al.* Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr. Biol.* **12,** 1317–1319 (2002).
72. Jones, S. J. *et al.* Changes in gene expression associated with developmental arrest and longevity in *Caenorhabditis elegans*. *Genome Res.* **11,** 1346–1352 (2001).
73. Hillier, L. W. *et al.* Genomics in *C. elegans*: so many genes, such a little worm. *Genome Res.* **15,** 1651–1660 (2005).
74. Shaw, W. M., Luo, S., Landis, J., Ashraf, J. & Murphy, C. T. The *C. elegans* TGF-β dauer pathway regulates longevity via insulin signaling. *Curr. Biol.* **17,** 1635–1645 (2007).
75. Hedgecock, E. M. & Herman, R. K. The *ncl-1* gene and genetic mosaics of *Caenorhabditis elegans*. *Genetics* **141,** 989–1006 (1995).

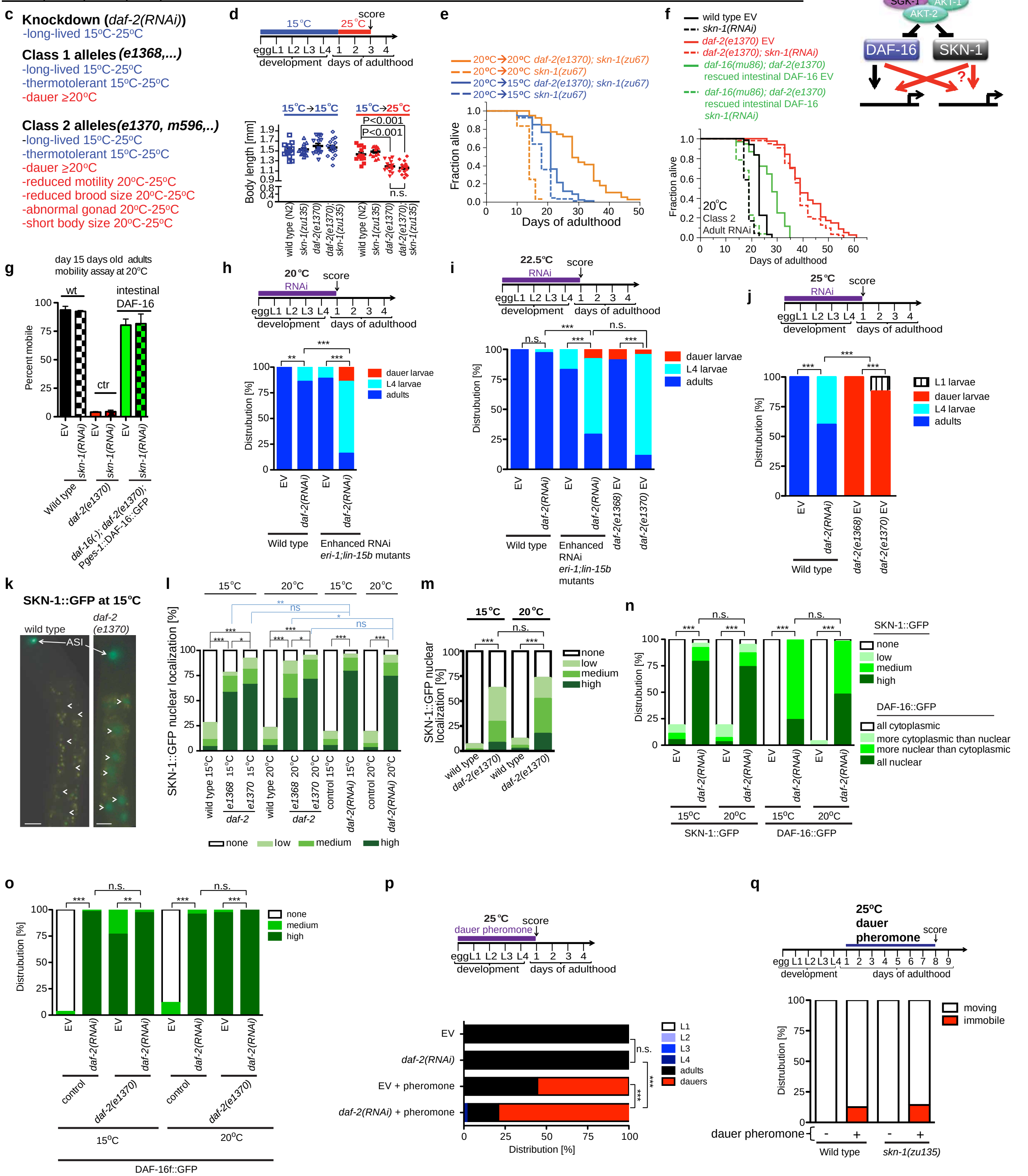
76. Schultz, R. D. & Gumienny, T. L. Visualization of *Caenorhabditis elegans* cuticular structures using the lipophilic vital dye DiI. *J. Vis. Exp.* **59,** e3362 (2012).
77. Kim, T. H., Kim, Y. J., Cho, J. W. & Shim, J. A novel zinc-carboxypeptidase SURO-1 regulates cuticle formation and body morphogenesis in *Caenorhabditis elegans. FEBS Lett.* **585,** 121–127 (2011).
78. Viswanathan, M., Kim, S. K., Berdichevsky, A. & Guarente, L. A role for SIR-2.1 regulation of ER stress response genes in determining *C. elegans* life span. *Dev. Cell* **9,** 605–615 (2005).
79. Menzel, R. *et al.* The nematode *Caenorhabditis elegans*, stress and aging: identifying the complex interplay of genetic pathways following the treatment with humic substances. *Front. Genet.* **3,** 50 (2012).
80. Pietsch, K. *et al.* Meta-analysis of global transcriptomics suggests that conserved genetic pathways are responsible for quercetin and tannic acid mediated longevity in *C. elegans. Front. Genet.* **3,** 48 (2012).
81. Gusarov, I. *et al.* Bacterial nitric oxide extends the lifespan of *C. elegans. Cell* **152,** 818–830 (2013).
82. Schmeisser, S. *et al.* Neuronal ROS signaling rather than AMPK/sirtuin-mediated energy sensing links dietary restriction to lifespan extension. *Mol. Metabol.* **2,** 1–11 (2013).
83. Staab, T. A. *et al.* The conserved SKN-1/Nrf2 stress response pathway regulates synaptic function in *Caenorhabditis elegans. PLoS Genet.* **9,** e1003354 (2013).
84. Iser, W. B., Wilson, M. A., Wood, W. H., III, Becker, K. & Wolkow, C. A. Co-regulation of the DAF-16 target gene, *cyp-35B1/dod-13*, by HSF-1 in *C. elegans* dauer larvae and *daf-2* insulin pathway mutants. *PLoS ONE* **6,** e17369 (2011).
85. Zarse, K. *et al.* Impaired insulin/IGF1 signaling extends life span by promoting mitochondrial L-proline catabolism to induce a transient ROS signal. *Cell Metab.* **15,** 451–465 (2012).
86. Halaschek-Wiener, J. *et al.* Analysis of long-lived *C. elegans daf-2* mutants using serial analysis of gene expression. *Genome Res.* **15,** 603–615 (2005).
87. Mair, W. *et al.* Lifespan extension induced by AMPK and calcineurin is mediated by CRTC-1 and CREB. *Nature* **470,** 404–408 (2011).
88. Greer, E. L. *et al.* Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in *C. elegans. Nature* **466,** 383–387 (2010).
89. Cristina, D., Cary, M., Lunceford, A., Clarke, C. & Kenyon, C. A regulated response to impaired respiration slows behavioral rates and increases lifespan in *Caenorhabditis elegans. PLoS Genet.* **5,** e1000450 (2009).
90. Chen, D. *et al.* Germline signaling mediates the synergistically prolonged longevity produced by double mutations in *daf-2* and *rsks-1* in *C. elegans*. *Cell Rep.* **5,** 1600–1610 (2013).
91. Chen, S. *et al.* The conserved NAD(H)-dependent corepressor CTBP-1 regulates *Caenorhabditis elegans* life span. *Proc. Natl Acad. Sci. USA* **106,** 1496–1501 (2009).
92. Huang da. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13 (2009).
93. Huang, da. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).

a

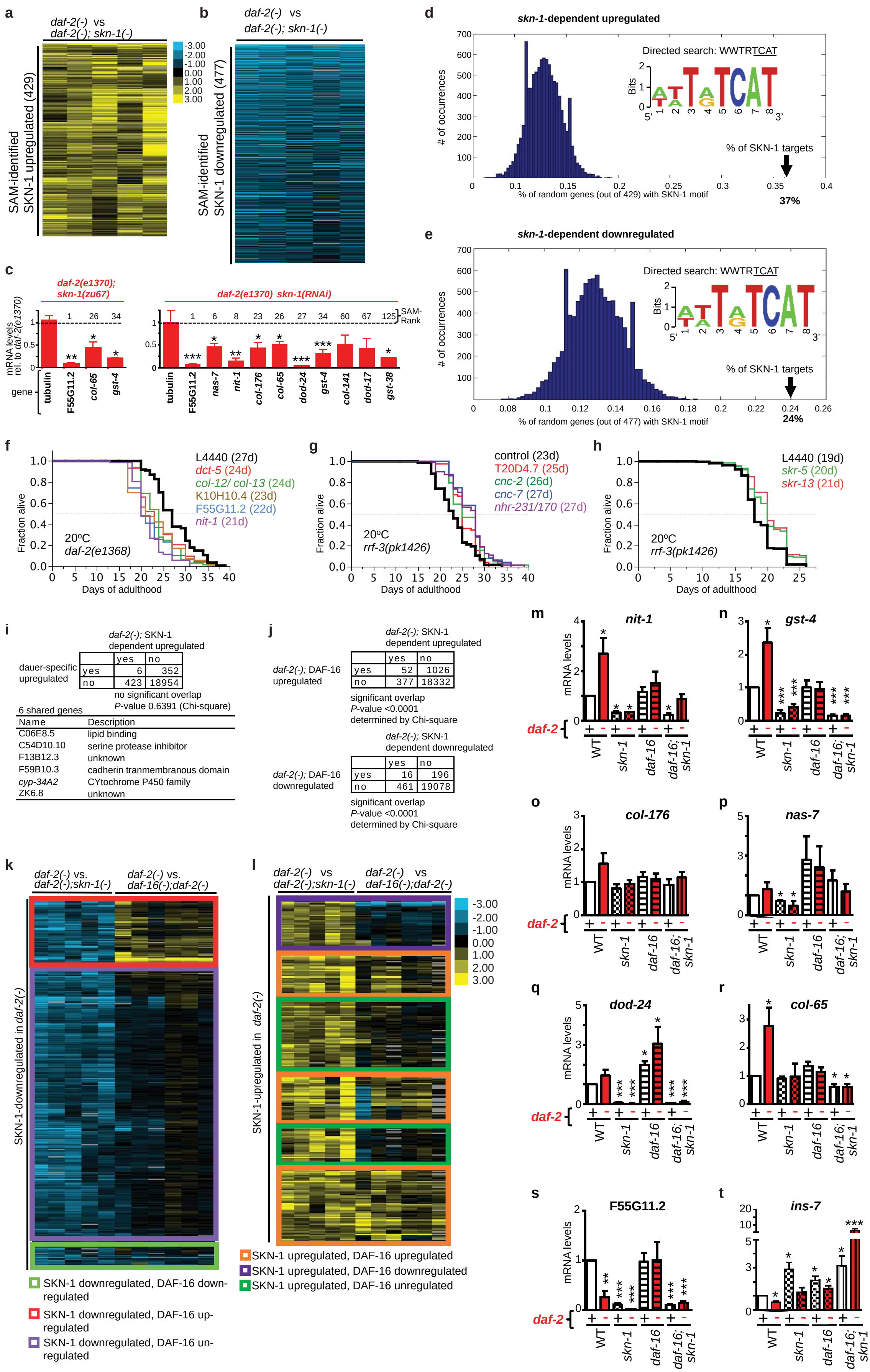
Genotype	Lifespan extension 15°C	*skn-1*-dependent 15°C	Lifespan extension 20°C	*skn-1*-dependent 20°C	Lifespan extension 25°C	*skn-1*-dependent 25°C	*Dauer entry 25°C	*Class 2 traits 20°C
daf-2(RNAi)	56%	Yes	62%	Yes	64%	Yes	No	No
daf-2(e1368) Class 1	34%	Yes	46%	Partial	ND	ND	Yes	No
daf-2(e1370) Class 2	50%	Yes	53%	No	ND	ND	Yes	Yes
daf-2(e1368) daf-2(RNAi)	193%	Yes	139%	Partial	ND	ND	Yes	No

b



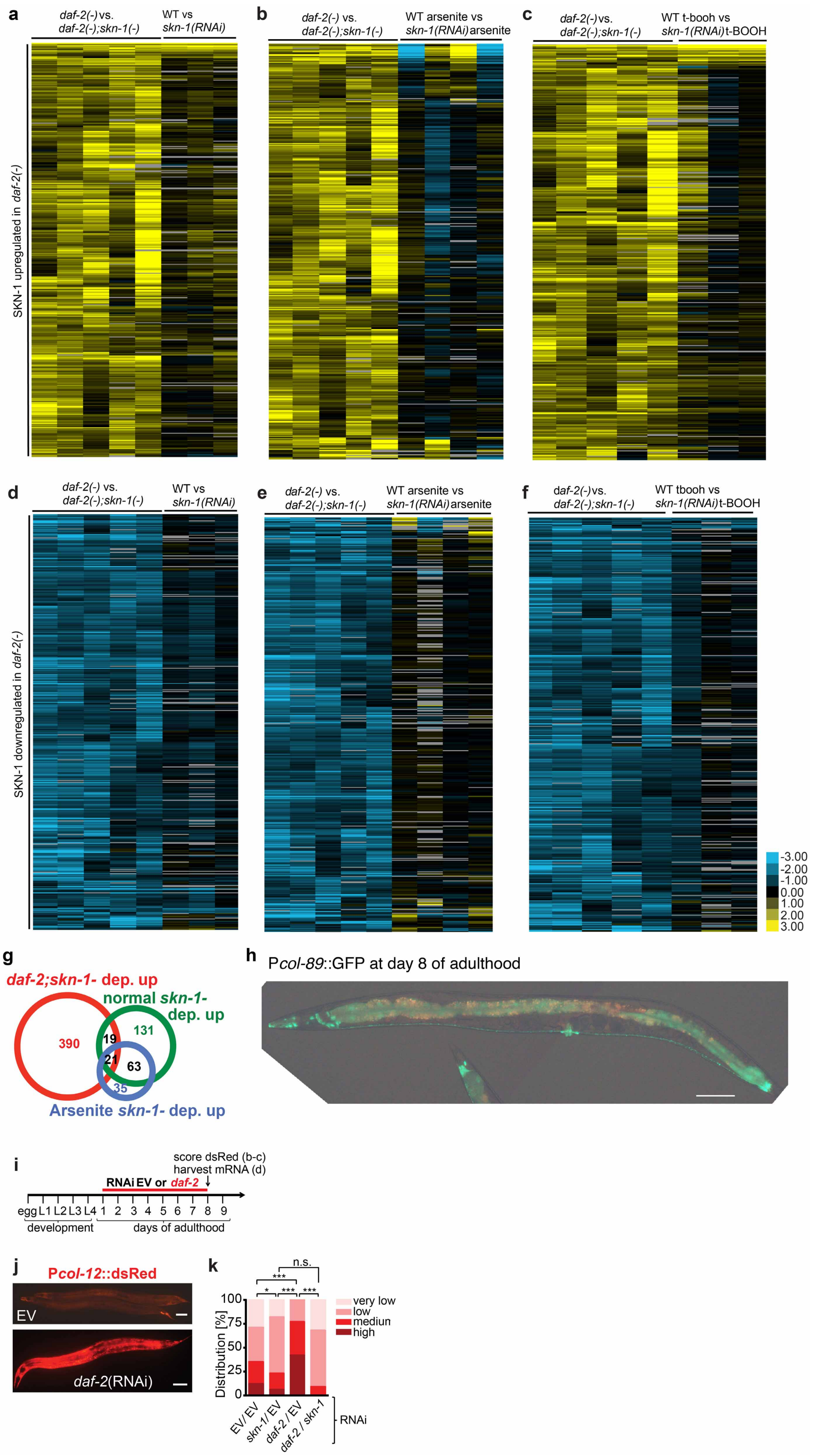


Extended Data Figure 1 | Analyses of rIIS under dauer-independent and dauer-predisposed conditions. **a**, Data from this study illustrating that rIIS longevity dependence upon *skn-1* correlates with low dauer pathway activity, not temperature or percentage increase in mean lifespan extension (*described in the Supplementary Discussion). **b**, Partial schematic of the IIS pathway in *C. elegans*. Insulin-like peptides (ins) bind to DAF-2, leading to activation of the AKT-1/2 and possibly SGK-1 kinases[1,13,63], which phosphorylate DAF-16 and SKN-1. Class 1 *daf-2* mutations are typically located on the extracellular portion of DAF-2, whereas most class 2 mutations affect its intracellular domains[64]. **c**, Mutant phenotypes of *daf-2*. Red indicates penetrance specifically at higher temperatures (Supplementary Discussion). **d**, The class 2 (dauer-related) *daf-2* trait of reduced body length is *skn-1*-independent. Each dot represents an animal, with *P* values determined by one-way analysis of variance (ANOVA) with post hoc Tukey's test. **e**, Dependence of dauer-independent *daf-2* longevity on adulthood *skn-1*. *daf-2*(*e1370*) lifespan extension requires *skn-1* when the temperature is downshifted to 15 °C specifically during adulthood (blue). For additional information see Supplementary Table 2. **f**, The *skn-1* dependence of *daf-2*(*e1370*) longevity at 20 °C when DAF-16 is expressed specifically in the intestine (strain description in Extended Data Table 1). **g**, Intestine-specific DAF-16 expression fails to rescue a class 2 dauer-like trait (immobility) in *daf-2*(*e1370*). **h**–**j**, Condition-specific induction of dauer by *daf-2* RNAi. The *daf-2* RNAi fails to induce dauer entry even at 25 °C (**j**), although some dauers are seen under more extreme conditions (27 °C)[65]. The activity of IIS and DAF-16 in neurons is critical for dauer regulation[15,16,66], and in the wild-type RNAi is comparatively ineffective in neurons[67], suggesting that the extremely weak dauer propensity of *daf-2* RNAi might derive from a failure to reduce IIS sufficiently in neurons. Supporting this idea, *daf-2* RNAi induced dauer entry even at 20 °C in *eri-1*(*mg366*)*; lin-15B*(*n744*) mutants, in which neuronal RNAi is robust[68] (**h**). $N > 100$ for each condition, two merged trials. **k**–**n**, Robust SKN-1 and DAF-16 nuclear localization under conditions of dauer inactivity. SKN-1 nuclear accumulation is inhibited comparably by IIS at 15 and 20 °C. SKN-1 is constitutively localized to ASI neuron nuclei in wild-type animals, and accumulates in intestinal nuclei in *daf-2*(*e1370*)[13]. **k**, Extent of IIS reduction from *daf-2*(*e1370*) at 15 °C, indicated by nuclear SKN-1::GFP. Chevrons indicate intestinal nuclei; scale bar, 20 µm. SKN-1::GFP (LD001) in intestinal nuclei is quantified in **l**, **m**. $N > 60$ for each condition and trial, three merged trials with *P* values determined by χ^2 test. Nuclear accumulation was scored as in Methods. **n**, The *daf-2* RNAi comparably induces SKN-1::GFP (LD001) and DAF-16::GFP (TJ356) intestinal nuclear localization at 15 and 20 °C. $N > 60$ for each condition, one trial with all experimental conditions done in parallel. **o**, Comparable nuclear accumulation of DAF-16f::GFP (*lpIs14*) induced by *daf-2* RNAi and *daf-2*(*e1370*) at 15 and 20 °C. $N > 60$ for each condition, one trial performed in parallel. **p**, **q**, Induction of dauer development (**p**) and dauer-like traits (*skn-1*-independent) (**q**) by the crude dauer pheromone preparation used in lifespan assays (Fig. 1e, Extended Data Table 1 and Supplementary Table 2). In **p**, $N > 100$ for each condition, one trial. In **q**, $N = 30$ for each condition, three merged trials. For **h**–**j**, **l**–**o**, *P* values were determined by χ^2 test; n.s., not significant, *$P < 0.05$, **$P < 0.001$, ***$P < 0.0001$.

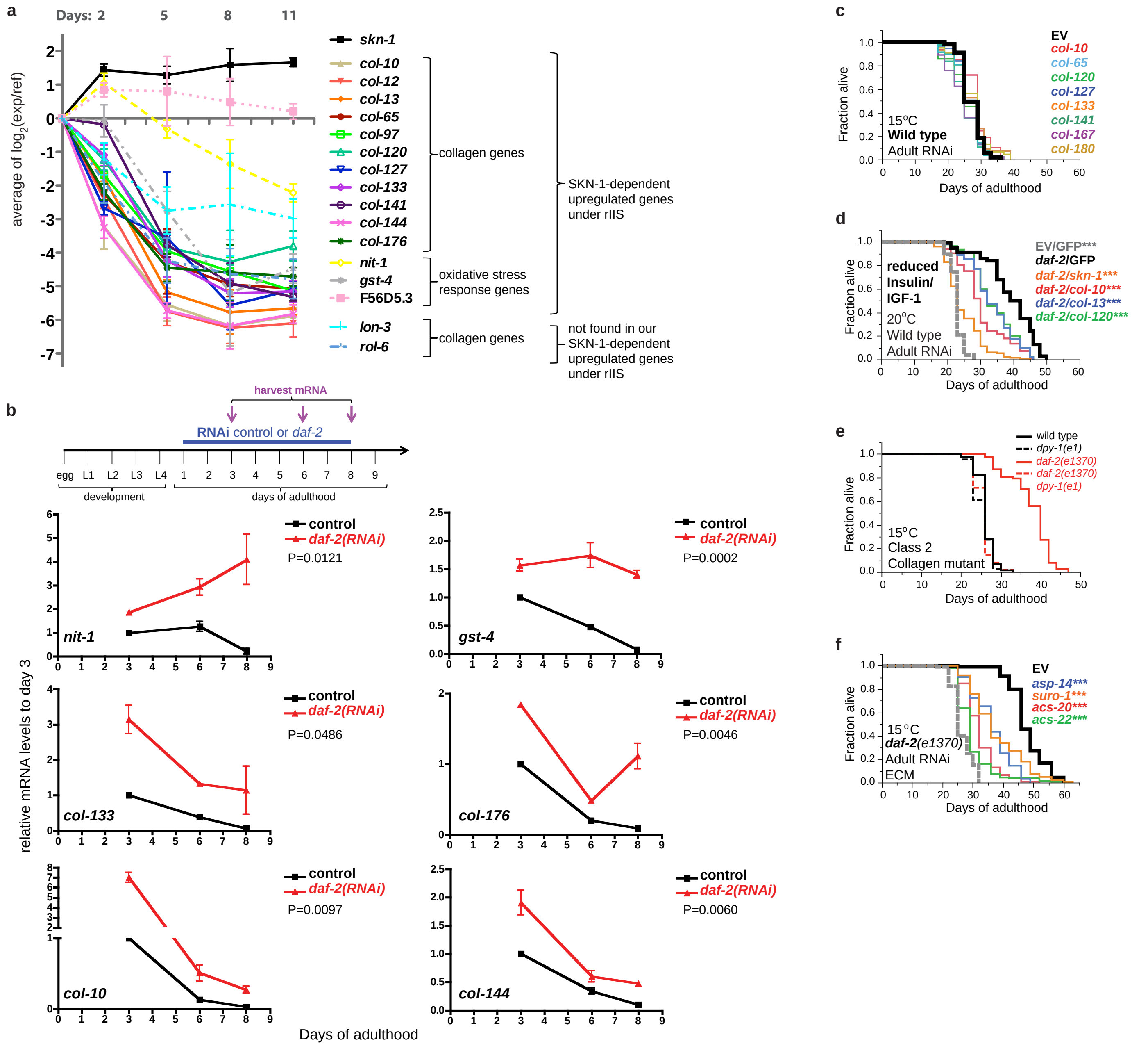


a
daf-2(-) vs daf-2(-); skn-1(-)
SAM-identified SKN-1 upregulated (429)
-3.00 -2.00 -1.00 0.00 1.00 2.00 3.00
b
daf-2(-) vs daf-2(-); skn-1(-)
SAM-identified SKN-1 downregulated (477)
c
daf-2(e1370); skn-1(zu67)
daf-2(e1370) skn-1(RNAi)
mRNA levels rel. to daf-2(e1370)
SAM-Rank
gene
tubulin F55G11.2 col-65 gst-4
tubulin F55G11.2 nas-7 nit-1 col-176 col-65 dod-24 gst-4 col-141 dod-17 gst-38
d
skn-1-dependent upregulated
Directed search: WWTRTCAT
# of occurrences
% of SKN-1 targets
% of random genes (out of 429) with SKN-1 motif
37%
e
skn-1-dependent downregulated
Directed search: WWTRTCAT
% of random genes (out of 477) with SKN-1 motif
24%
f
L4440 (27d)
dct-5 (24d)
col-12/ col-13 (24d)
K10H10.4 (23d)
F55G11.2 (22d)
nit-1 (21d)
20ºC
daf-2(e1368)
Fraction alive
Days of adulthood
g
control (23d)
T20D4.7 (25d)
cnc-2 (26d)
cnc-7 (27d)
nhr-231/170 (27d)
20ºC
rrf-3(pk1426)
h
L4440 (19d)
skr-5 (20d)
skr-13 (21d)
20ºC
rrf-3(pk1426)
i
daf-2(-); SKN-1 dependent upregulated
dauer-specific upregulated
yes | no
yes 6 | 352
no 423 | 18954
no significant overlap
P-value 0.6391 (Chi-square)
6 shared genes
Name | Description
C06E8.5 | lipid binding
C54D10.10 | serine protease inhibitor
F13B12.3 | unknown
F59B10.3 | cadherin tranmembranous domain
cyp-34A2 | CYtochrome P450 family
ZK6.8 | unknown
j
daf-2(-); SKN-1 dependent upregulated
daf-2(-); DAF-16 upregulated
yes | no
yes 52 | 1026
no 377 | 18332
significant overlap
P-value <0.0001
determined by Chi-square
daf-2(-); SKN-1 dependent downregulated
daf-2(-); DAF-16 downregulated
yes | no
yes 16 | 196
no 461 | 19078
significant overlap
P-value <0.0001
determined by Chi-square
k
daf-2(-) vs. daf-2(-);skn-1(-)
daf-2(-) vs. daf-16(-);daf-2(-)
SKN-1-downregulated in daf-2(-)
SKN-1 downregulated, DAF-16 down-regulated
SKN-1 downregulated, DAF-16 up-regulated
SKN-1 downregulated, DAF-16 un-regulated
l
daf-2(-) vs daf-2(-);skn-1(-)
daf-2(-) vs daf-16(-);daf-2(-)
SKN-1-upregulated in daf-2(-)
SKN-1 upregulated, DAF-16 upregulated
SKN-1 upregulated, DAF-16 downregulated
SKN-1 upregulated, DAF-16 unregulated
m nit-1
n gst-4
o col-176
p nas-7
q dod-24
r col-65
s F55G11.2
t ins-7
mRNA levels
daf-2
WT skn-1 daf-16 daf-16; skn-1

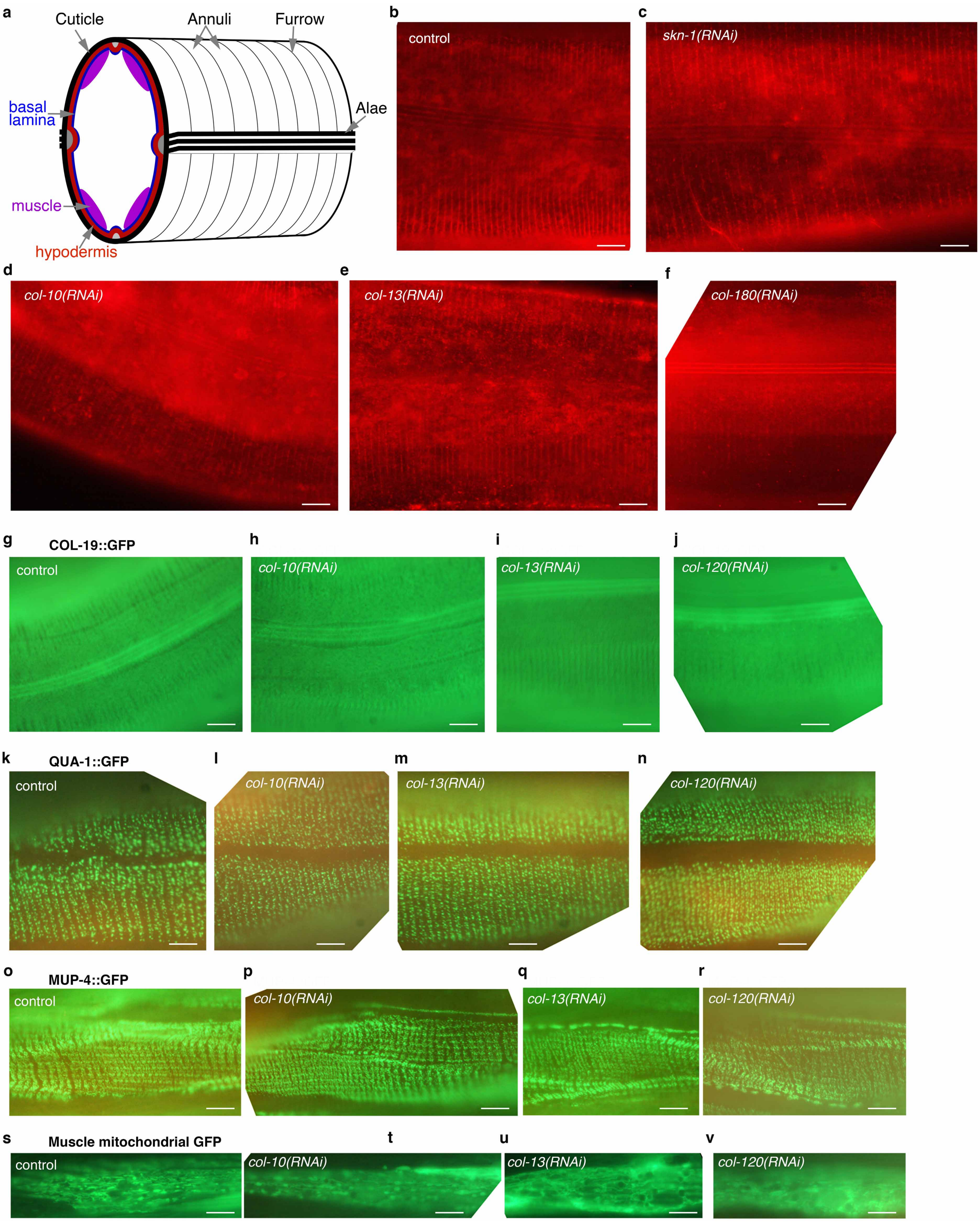
Extended Data Figure 2 | Identification of SKN-1-regulated *daf-2*(−) genes. **a**, Heatmap of 429 genes identified by SAM as significantly upregulated by SKN-1 in *daf-2* mutants. **b**, Four hundred and seventy-seven genes identified by SAM as significantly downregulated by SKN-1 in *daf-2* mutants (Supplementary Table 3). The SKN-1-downregulated *daf-2*(−) set was enriched for genes involved in ubiquitin-mediated proteolysis (E3 ligase/SCF, F-box; Supplementary Table 8). Columns represent biological samples. Blue, down; black, unregulated. **c**, Confirmation of microarray data for SKN-1-upregulated *daf-2*(−) genes by qPCR at 15 °C. One and three biological replicates were analysed in the left and right panels, respectively. SAM scores are in Supplementary Table 3. Data are mean ± s.e.m. $*P < 0.05$, $**P < 0.001$, $***P < 0.0001$ relative to *daf-2*, determined by one sample *t*-test, two-tailed, hypothetical mean of 1. **d**, **e**, Enrichment of SKN-1 binding sites upstream of SKN-1-regulated *daf-2*(−) genes. An unbiased search using the Weeder and FIRE algorithms did not detect any overrepresented form of the consensus SKN-1 binding motif (WWTRTCAT) (W = A/T, R = G/A)[69]. Given the degeneracy of this motif, we used RSATools to perform a directed search of 600 bp upstream of SKN-1 upregulated (**d**) and downregulated (**e**) genes. This search window was based upon the location of SKN-1 binding sites identified by genome-wide chromatin immunoprecipitation followed by sequencing (ChIP-seq) using transgenically expressed SKN-1 (ref. 70). A SKN-1 motif was detected at only 13% of a random sample of 10,000 genes, but at 37% and 24% of the SKN-1-upregulated (out of 429 genes) and downregulated genes (out of 477 genes), respectively. **f**, Importance of SKN-1-upregulated *daf-2*(−) genes for *daf-2*(*e1368*) lifespan. The class 1 *daf-2* allele *e1368* is partly dependent upon *skn-1* for lifespan extension at 20 °C (Extended Data Table 1)[13]. Adult RNAi against 5 of 12 genes tested reduced *daf-2*(*e1368*) lifespan at 20 °C. **g**, **h**, Several SKN-1-downregulated *daf-2*(−) genes decrease lifespan. Knockdown was performed in the RNAi-sensitive strain *rrf-3*(*pk1426*)[71]. **g**, Genes for which RNAi knockdown increased lifespan, from 12 that were analysed without regard to their function. **h**, Analysis of six Skp1-related genes, an overrepresented category among SKN-1-downregulated *daf-2*(−) genes (Supplementary Table 8). Only genes that affected lifespan are shown. Other data and all statistics are in Supplementary Table 6. For 15 other SKN-1-downregulated *daf-2*(−) genes, it has been shown previously that RNAi increases lifespan (Supplementary Table 5). Parts **f** and **g** each show a single trial, and a composite of three trials is shown in **h**. In **g** the negative RNAi control is *elpc-4*(*RNAi*) instead of L4440. Mean lifespan (in days) is indicated for each gene. **i**, Overlap between the *daf-2*(−); SKN-1-dependent upregulated gene set (429 genes, this study) and a set of genes preferentially upregulated in dauers (358 genes)[72]. The overlap of six genes was not significant ($P = 0.6391$ by two-sided χ^2 test). The number of genes that were present in neither set (no/no) was determined by subtracting the total number in both gene sets from the total number of genes encoded in *C. elegans* 19,735 (ref. 73). **j**, Overlaps between SKN-1-regulated *daf-2*(−) and DAF-16-regulated *daf-2*(−) gene sets[74]. For both up- and downregulated *daf-2*(−) genes, overlaps between the SKN-1- and DAF-16-regulated sets were significant ($P < 0.0001$ determined by two-sided χ^2 test). Moreover, hierarchical clustering identified additional SKN-1-upregulated *daf-2*(−) genes that were also upregulated by DAF-16 even though they were not present in this list of highest-confidence DAF-16-regulated genes (**l**). The number of genes that were in neither set (no/no) was determined as in **i**. **k**, Hierarchical clustering of SKN-1-downregulated *daf-2*(−) gene sets with DAF-16-regulated genes. SKN-1-regulated genes identified here were queried as to how they were influenced by DAF-16 in a comparison of *daf-2*(*e1370*) versus *daf-16*(*mu86*)*; daf-2*(*e1370*) animals raised at 20 °C (ref. 74). Three hundred and ninety-three SKN-1-upregulated *daf-2*(−) genes that were present in this DAF-16-regulated data set are shown. Most SKN-1-downregulated *daf-2*(−) genes did not appear to be regulated by DAF-16. **l**, Hierarchical clustering of SKN-1-upregulated *daf-2*(−) genes with DAF-16-regulated genes that were identified by comparing *daf-2*(*e1370*) versus *daf-16*(*mu86*)*; daf-2*(*e1370*) at 20 °C (ref. 74). Two hundred and seventy-two SKN-1-upregulated *daf-2*(−) genes that were present in this DAF-16-regulated data set are shown, 46% of which were upregulated by both SKN-1 and DAF-16. Yellow, up; blue, down; black, unregulated. **m–t**, Effects of SKN-1 and DAF-16 on individual genes in response to *daf-2* RNAi at 20 °C. A qPCR analysis of *skn-1*(*zu67*), *daf-16*(*mgDf47*), and *daf-16*(*mgDf47*)*;skn-1*(*zu67*) double mutants indicated that many genes are upregulated by *daf-2*(*RNAi*) (red) in a *skn-1*-dependent manner, but also that these genes vary in how they are affected by DAF-16. DAF-16 and SKN-1 increased activity of *gst-4*, *col-65*, and *col-176*, but DAF-16 seemed to downregulate *dod-24*, *nas-7*, and *F55G11.2*. All of these genes except *ins-7* were identified in our *daf-2; skn-1* data sets. For each condition, three biological samples of 200 worms each were analysed by qPCR. All data are mean ± s.e.m. $*P < 0.05$, $**P < 0.001$, $***P < 0.0001$ relative to wild-type RNAi control, determined by one sample *t*-test, two-tailed, hypothetical mean of 1.



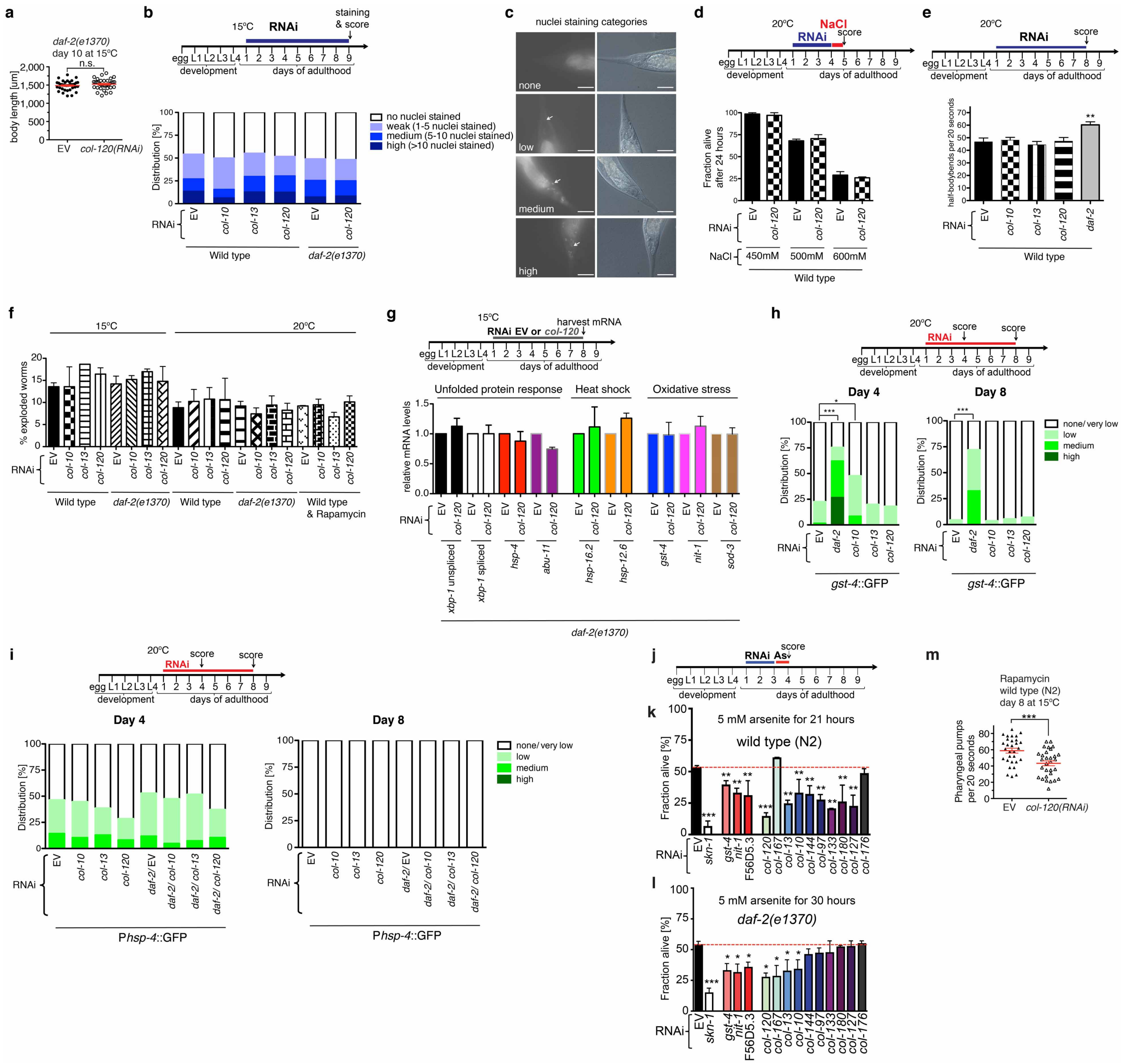
Extended Data Figure 3 | Analyses of SKN-1-regulated *daf-2*(−) genes. **a**–**f**, SKN-1-upregulated (**a**–**c**) and downregulated (**d**–**f**) *daf-2*(−) gene sets were examined by hierarchical clustering to determine how they were previously found to be affected by SKN-1 under unstressed or oxidative stress conditions[18]. t-BOOH, *tert*-butyl hydroperoxide. **g**, Proportional Venn diagrams show comparisons of SKN-1-upregulated genes identified under *daf-2*(−), normal, or arsenite treatment conditions[18] (Supplementary Table 7). In each case, L4 larvae were analysed to avoid embryogenesis effects. Heatmaps are shown in **a**–**f**. **h**, The SKN-1-upregulated *daf-2*(−) collagen *col-89* is expressed in neurons and the intestine, but not in hypodermis. Transgenic P*col-89*::GFP (BC12533) at day 8 of adulthood is shown. Anterior to the left, ventral side down; scale bar, 100 μm. **i**–**k**, SKN-1-mediated collagen gene activation in day 8 *daf-2*(*RNAi*) adults. Adulthood *daf-2* knockdown (**i**) activated a P*col-12*::dsRED reporter (**j**; scale bar, 100 μm). **k**, The *skn-1* dependence of P*col-12*::dsRED expression. EV, empty RNAi vector. $N > 60$ for each condition, three merged trials, with P value by χ^2 test (*$P < 0.05$; ***$P < 0.0001$; n.s., not significant).



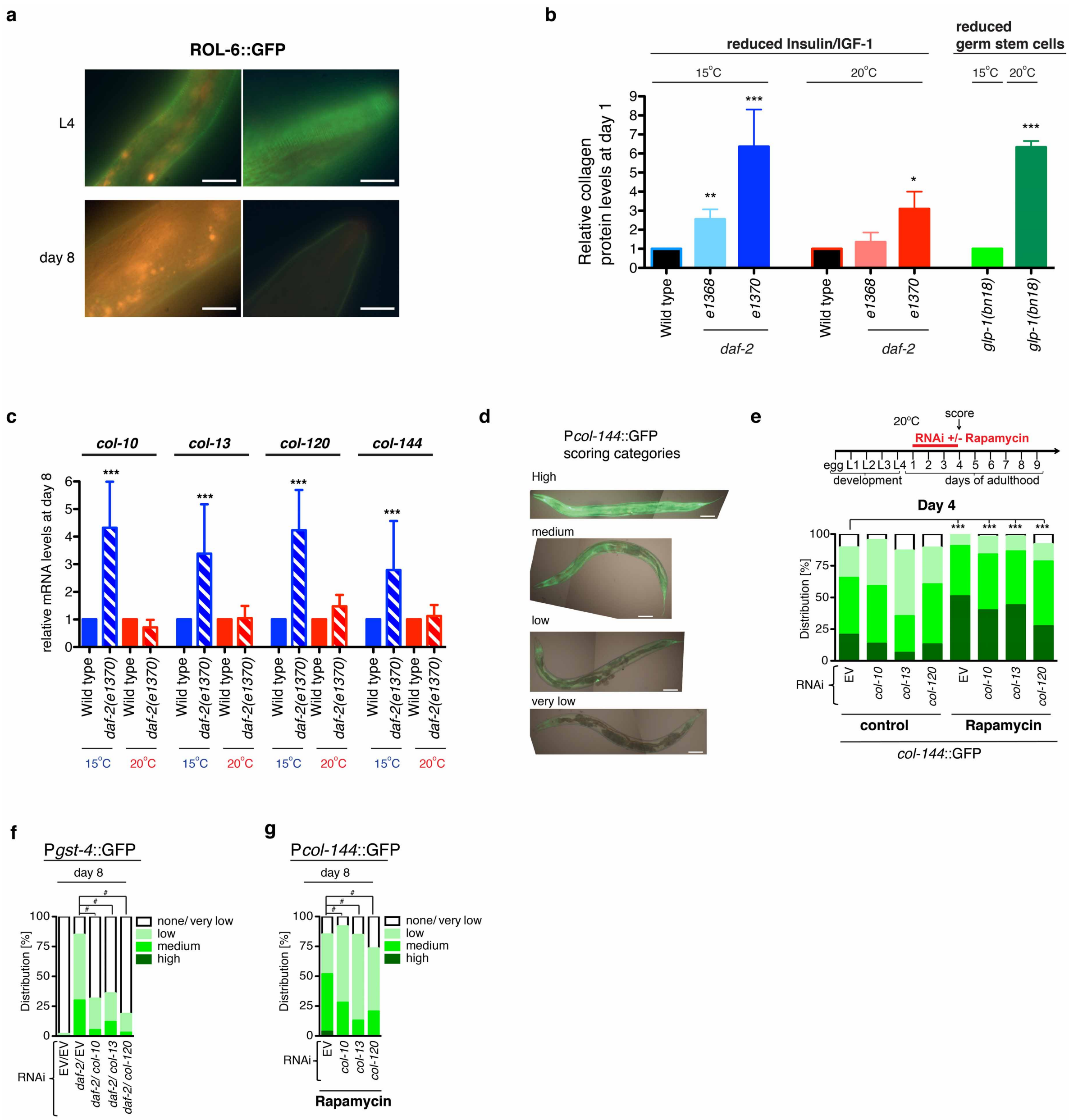
Extended Data Figure 4 | rIIS delays age-associated decline in collagen expression. **a**, Age-associated decline in expression of selected collagen and SKN-1-dependent detoxification genes. Eighty-eight collagens are among many genes that decline in expression as *C. elegans* ages[22]. Fifty of these age-downregulated genes were in our SKN-1 upregulated *daf-2*(−) gene set, including 27 collagen genes (Supplementary Table 10). These *daf-2*(−); SKN-1-dependent collagens were neither flanked by SKN-1 binding sites nor bound by SKN-1 in a genome-wide survey (Supplementary Table 9)[70], suggesting that they are regulated by SKN-1 indirectly. The average Cy5-labelled cDNA values of day 2–11 adults (indicated as 'exp') are plotted in binary logarithm ($\log_2$) relative to cy3-labelled reference cDNA from mixed stage hermaphrodites (indicated as 'ref'). Data are from ref. 22. The *nit-1*, *gst-4*, and *F56D5.3* genes are predicted to encode a nitrilase, glutathione *S*-transferase, and NADPH oxidoreductase, respectively (WormBase). **b**, Expression of SKN-1-regulated collagen and oxidative stress response genes (*nit-1* and *gst-4*) are maintained during ageing in *daf-2*(*RNAi*) animals. One-day-old adult wild-type (N2) animals were placed on either empty vector control (L4440) (black) or *daf-2* RNAi (red) at 20 °C. mRNA was harvested at days 3, 6, and 8. mRNA levels are shown relative to wild-type (N2) day 3 adults on empty vector control (L4440) RNAi and are represented as mean ± s.e.m. For each condition, two biological samples of more than 100 worms each were analysed by qPCR. For each gene, the statistical difference of relative mRNA expression levels between L4440 and *daf-2*(*RNAi*) treatment over the time course (days 3, 6, 8) is shown by two-way ANOVA (repeated measures). **c**, Adulthood knockdown of SKN-1-upregulated collagens did not affect wild-type lifespan For statistics and additional trials, see Extended Data Table 3 and Supplementary Table 13. **d**, Importance of SKN-1-upregulated collagens for *daf-2*(*RNAi*) longevity. Adulthood RNAi knockdown of *daf-2* combined with collagens or *skn-1* at 20 °C is shown. GFP was the RNAi control. For statistics and additional data, see Supplementary Table 13. **e**, Suppression of *daf-2*(*e1370*) but not wild-type longevity at 15 °C by the collagen mutation *dpy-1*(*e1*), which affects the cuticle[31,75], but was not present in our SKN-1-regulated gene set. For details and statistics see Supplementary Table 13. **f**, Longevity of *daf-2*(*e1370*) at 15 °C requires the SKN-1-upregulated extracellular proteases *asp-14* and *suro-1*, along with cuticle integrity genes *acs-20* and *acs-22* (ref. 62), suggesting a general importance of ECM gene expression. For details and statistics see Supplementary Table 13.



Extended Data Figure 5 | Adulthood knockdown of collagens important for longevity does not affect morphology of cuticle-associated structures. **a**, Schematic cross-section of *C. elegans* illustrating the proximity of the cuticle (black), hypodermis (red), basal lamina (blue), and body-wall muscles (purple). Annuli, furrow, and alae are characteristic cuticle structures. **b**–**j**, Adulthood RNAi against SKN-1-upregulated *daf-2*(−) collagens does not affect cuticle morphology. **b**–**f**, One-day-old wild-type animals were exposed to either empty vector (control) or the indicated RNAi clone by feeding. Ten days later, animals were incubated in DiI for 16 h; the cuticle was imaged as described in ref. 76. $N > 30$ animals per condition scored, with typical images shown. Scale bar, 10 μm. **g**–**j**, Cuticle morphology revealed by the collagen COL-19, detected by a translational fusion protein (*kaIs12* [COL-19::GFP]). We did not identify *col-19* as being regulated by *daf-2* and *skn-1*, and *daf-2*(*RNAi*) did not detectably alter COL-19::GFP levels (not shown). **k**–**n**, Adulthood knockdown of SKN-1-upregulated *daf-2*(−) collagens does not affect the pattern of *chEx1682* QUA-1::GFP, a marker of cuticle adhesion. QUA-1 encodes a hedgehog-related protein required for moulting, cuticle adhesion, and alae formation[42]. **o**–**r**, Adulthood RNAi against SKN-1-upregulated *daf-2*(−) collagens does not affect the pattern of muscle–hypodermis–cuticle adhesion, as indicated by *upIs1* MUP-4::GFP. MUP-4 is a transmembrane protein that is part of a complex that attaches hypodermis and muscles to the cuticle[35]. **s**–**v**, Adulthood collagen knockdown does not affect mitochondrial morphology in muscle. For **g**–**v**, animals were placed on RNAi at the first day of adulthood and scored and imaged at day 8 of adulthood. $N > 30$ animals per condition scored, with typical images shown. Scale bar, 10 μm.



Extended Data Figure 6 | Phenotypic analyses of collagens important for longevity. **a**, Adulthood *col-120* knockdown does not affect *daf-2*(*e1370*) body size at 15 °C. The *daf-2*(*e1370*) animals were placed on RNAi food as day 1 adults, and at day 10 body size, pharyngeal pumping, and lipofuscin levels were scored in parallel in the same animals ($N > 30$; one trial; see Fig. 4a, b). **b**, **c**, Adulthood knockdown of SKN-1-upregulated collagens does not alter barrier function. **b**, Upper panel: animals were placed on RNAi food on adulthood day 1, and at day 9 were incubated in 1 μg ml^{-1} Hoechst 33342, which is membrane-permeable but cuticle-impermeable. For details see Methods, adapted from ref. 62. Lower panel: barrier permeability was not affected by *daf-2* mutation or collagen knockdown. Permeability was assessed by nuclear Hoechst staining in the tail[62] ($N > 50$ per condition; one trial). Approximately half of the animals in each group showed nuclear staining in the tail that is likely to have arisen through uptake in the intestine, as suggested by the high levels of intestinal Hoechst staining (**c**). Uptake through the cuticle would have resulted in a much wider distribution of stained nuclei. **c**, Representative pictures of quantification categories. Arrow indicates Hoechst-stained tail nuclei. Scale bar, 50 μm. **d**, Adulthood knockdown of *col-120* did not sensitize to hypertonic stress. Day 1 adult wild-type animals were placed RNAi food for 3 days, then on plates containing food and high concentrations of salt for 24 h before assay (NaCl: 450 mM, 500 mM, 600 mM; $N > 60$ per condition; two trials). **e**, Adulthood knockdown of SKN-1-upregulated collagens did not impair body movement. Neither the frequency nor morphology (not shown) of body movement was affected. In parallel, the *daf-2* RNAi control increased movement frequency because these animals were chronologically younger. (**$P < 0.001$, one-way ANOVA post hoc Tukey's test compared with empty RNAi vector.) **f**, Adulthood collagen RNAi did not increase vulval rupturing during ageing. The bar graph shows the mean ± s.e.m. percentage of exploded worms that were censored during lifespan assays (Extended Data Table 3 and Supplementary Table 13). **g**–**i**, Adulthood *col-120* knockdown did not induce unfolded protein, heat-shock stress, or oxidative stress responses. In **g**, *daf-2*(*e1370*) mutants were placed on RNAi food as day 1 adults, and assayed at day 8 (upper panel). Relative levels of these stress response gene mRNAs were determined by qPCR (two independent trials, each with 200 worms per condition). **h**, Adulthood collagen RNAi does not activate the oxidative stress response marker P*gst-4*::GFP[34], assayed after 4 and 8 days of RNAi. As a control, *daf-2* RNAi induced SKN-1 to increase *gst-4* expression (Fig. 2a). **i**, Adulthood collagen RNAi does not activate the unfolded protein response marker P*hsp-4*::GFP[40]. **j**–**l**, Importance of collagens for oxidative stress resistance. Day 1 adults were exposed to empty vector (EV) or RNAi food at 15 °C, then at day 3 were placed in 5 mM arsenite (As) and scored for survival. Knockdown of collagens and other SKN-1 upregulated *daf-2*(−) genes sensitized to oxidative stress from arsenite; *nit-1* (nitrilase), *gst-4* (glutathione *S*-transferase), *F56D5.3* (NADPH oxidoreductase). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ relative to control (empty RNAi vector), determined by one-way ANOVA with post hoc Tukey's test. t-BOOH experiments are described in Supplementary Table 16. **m**, Adulthood collagen expression required for rapamycin to delay appearance of an ageing marker (pharyngeal pumping). $N > 30$; each dot represents an animal; ***$P < 0.0001$ determined with an unpaired *t*-test, two-tailed.



Extended Data Figure 7 | Cuticle remodelling in adults a, The collagen ROL-6 is present in the cuticle during development and early adulthood[77], then largely disappears during ageing. The upper panels show the mid-body (left) and head (right) regions in an L4 animal. Day 1 adults exhibited similar levels and patterns of *jgIs5* ROL-6::GFP fluorescence (not shown). Lower panels show the corresponding regions in a day 8 adult, in which *jgIs5* ROL-6::GFP levels are reduced. The orange signal corresponds to gut autofluorescence. Representative images are shown; scale bar, 20 μm. $N = 30$ for each sample set (L4, day 1, and day 8). **b**, Total collagen levels are elevated in long-lived animals at the first day of adulthood. Note that these long-lived animals also maintain higher collagen levels in later life despite an age-related decline (Fig. 4d). Relative collagen levels were estimated by a hydroxyproline assay[61]. In *daf-2* mutants, total collagen levels were elevated at both temperatures but the increase was greater at 15 °C, at which *skn-1* and SKN-1-dependent collagens are required for lifespan extension (Fig. 3 and Supplementary Table 13). Temperature-sensitive *glp-1*(*bn18*) mutants were maintained at 15 °C (permissive temperature), or upshifted to 25 °C (restrictive temperature) as L1 larvae until the L4 stage, then placed at 20 °C. **c**, SKN-1-dependent collagen genes from the *daf-2*(−) set are not upregulated in 8-day-old *daf-2*(*e1370*) adults at 20 °C. Expression of these collagens remains increased at this age in *daf-2*(*e1370*) at 15 °C or after *daf-2* RNAi at 20 °C (Fig. 2a and Extended Data Figs 2c and 4b), conditions in which the dauer pathway is inactive and lifespan extension is *skn-1* dependent (see text). Two hundred day-8 adults were assayed in each sample, with three merged independent trials shown. **d**, Scoring categories for the P*col-144*::GFP reporter are shown in (**e**, **g**; Fig. 4f; scale bar, 100 μm). **e**, Adulthood rapamycin treatment increases *col-144* promoter activity. Knockdown of *col-10*, *col-13*, or *col-120* did not reduce P*col-144*::GFP levels at day 4, but significantly decreased P*col-144*::GFP levels by day 8 (**g**). $N > 60$ for each condition, two merged trials, with P value by χ^2 test ($^{\#}P < 0.0001$ against untreated empty RNAi vector control animals). **f**, Dependence of the SKN-1 target gene *gst-4* on adulthood SKN-1-upregulated collagen expression in *daf-2*(*RNAi*) animals. Collagen or empty vector (EV) control RNAi was initiated at day 1 of adulthood at 20 °C, together with *daf-2* knockdown. **g**, Adulthood collagen RNAi decreases *col-144* promoter activity in rapamycin-treated animals. As is seen in *daf-2* mutants at 15 °C (Fig. 4f), activity of this rapamycin-activated promoter is unaffected by adulthood collagen RNAi at day 4 (**e**), but reduced at day 8. For **f** and **g**, $N > 60$ for each condition, two merged trials, with P value by χ^2 test ($^{\#}P < 0.0001$).

Extended Data Table 1 | The *skn-1* dependence of *daf-2* lifespan extension in the absence of dauer-related mechanisms

Strain	Temp [°C]	Mean lifespan ± S.E.M. [Days]	75th per-centile [Days]	N assayed / Initial N	% mean lifespan change to N2 or control	% mean lifespan change to *skn-1*	P-value (log-rank) vs. N2	P-value (log-rank) vs. *skn-1*	P-value (log-rank) vs. *daf-2*	Figure
3 merged trials at 15°C *										
wild type (N2)	15	23.4 ± 0.3	27	279/332						1a
skn-1(zu67)	15	16.9 ± 0.2	19	278/315	-28		<0.0001			1a
daf-2(e1370)	15	36.7 ± 0.5	44	372/396	+57	+117	<0.0001	<0.0001		1a
daf-2(e1370); skn-1(zu67)	15	17.1 ±0.2	19	308/327		+1	<0.0001	0.7993	<0.0001	1a
Trial at 15°C and 20°C										
wild type (N2) L4440(RNAi)	15	25.4 ± 0.5	27	74/83						
wild type (N2) *daf-2*(RNAi)	15	39.1 ± 1.2	42	64/73	+54	+146	<0.0001	<0.0001		
skn-1(tm3411) L4440(RNAi)	15	15.9 ± 0.4	16	116/136	-37		<0.0001		<0.0001	
skn-1(tm3411) daf-2(RNAi)	15	16.6 ± 0.7	16	84/103		+6	<0.0001	0.6424	<0.0001	
wild type (N2) L4440(RNAi)	**20**	23.6 ± 0.5	26	45/51						1b
wild type (N2) *daf-2*(RNAi)	**20**	34.8 ± 1.1	40	45/50	+47	+120	<0.0001	<0.0001		1b
skn-1(tm3411) L4440(RNAi)	**20**	15.8 ± 0.5	16	93/108	-33		<0.0001		<0.0001	1b
skn-1(tm3411) daf-2(RNAi)	**20**	17.1 ± 0.5	16	115/124		+8	<0.0001	0.0491	<0.0001	1b
Trial at 20°C										
wild type (N2) L4440(RNAi)	**20**	23.3 ± 0.2	23	67/74						
wild type (N2) *skn-1*(RNAi)	**20**	18.6 ± 0.2	19	96/101	-20		<0.0001			ED 1f
daf-16(mu86);daf-2(e1370) L4440(RNAi)	**20**	17.0 ± 0.2	19	73/83	-27	-9	<0.0001	<0.0001	<0.0001	ED 1f
daf-16(mu86);daf-2(e1370) skn-1(RNAi)	**20**	16.5 ± 0.2	19	78/89	-29	-11	<0.0001	<0.0001	<0.0001	
daf-2(e1370) L4440(RNAi)	**20**	41.7 ± 0.9	47	82/87	+79	+124	<0.0001	<0.0001		ED 1f
daf-2(e1370) skn-1(RNAi)	**20**	38.9 ± 0.8	42	96/102	+67	+109	<0.0001	<0.0001	0.0153	ED 1f
DAF-16 rescued in all tissues: *daf-16(mgDf50);daf-2(e1370); Ipls14* [P*daf-16*::DAF-16f::GFP] L4440(RNAi)	**20**	53.5 ± 1.0	61	77/83	+130	+188	<0.0001	<0.0001	<0.0001	
DAF-16 rescued in all tissues: *daf-16(mgDf50);daf-2(e1370); Ipls14* [P*daf-16*::DAF-16f::GFP] *skn-1*(RNAi)	**20**	50.8 ± 0.9	58	86/93	+118	+173	<0.0001	<0.0001	<0.0001	
DAF-16 rescued in neurons: *daf-16(mu86);daf-2(e1370); muEx169* [P*unc119*::GFP::DAF-16] L4440(RNAi)	**20**	16.4 ± 0.2	17	71/77	-30	-12	<0.0001	<0.0001	<0.0001	
DAF-16 rescued in neurons: *daf-16(mu86);daf-2(e1370); muEx169* [P*unc119*::GFP::DAF-16] *skn-1*(RNAi)	**20**	16.5 ± 0.2	17	92/97	-29	-11	<0.0001	<0.0001	<0.0001	
DAF-16 rescued in intestine: *daf-16(mu86);daf-2(e1370); muEx211* [P*ges-1*::GFP::DAF-16] L4440(RNAi)	**20**	27.5 ± 0.6	30	89/98	+18	+48	<0.0001	<0.0001	<0.0001	ED 1f
DAF-16 rescued in intestine: *daf-16(mu86);daf-2(e1370); muEx211* [P*ges-1*::GFP::DAF-16] *skn-1*(RNAi)	**20**	18.5 ± 0.3	19	86/93	-21	-0.5	<0.0001	0.2671	<0.0001	ED 1f
Trial at 15°C and 20°C										
wild type (N2) L4440(RNAi)	**15**	24.5 ± 0.6	28	68/76						1c
wild type (N2) *daf-2*(RNAi)	**15**	47.8 ± 1.3	57	81/94	+95	+148	<0.0001	<0.0001	<0.0001$	1c
skn-1(zu135) L4440(RNAi)	**15**	19.3 ± 0.9	21	61/73	-21		<0.0001		<0.0001$	1d
skn-1(zu135) daf-2(RNAi)	**15**	20.3 ± 1.1	26	68/77	-17	+5	0.0026	0.3805	<0.0001$	1d
daf-2(e1368) L4440(RNAi)	**15**	32.9 ± 0.9	40	65/72	+34	+70	<0.0001	<0.0001		1c
daf-2(e1368) daf-2(RNAi)	**15**	71.7 ± 2.3	82	70/78	+193	+272	<0.0001	<0.0001	<0.0001$	1c
daf-2(e1368); skn-1(zu135) L4440(RNAi)	**15**	19.9 ± 0.7	26	48/57	-19	+3	0.0002	0.4624	<0.0001$	1d
daf-2(e1368); skn-1(zu135) daf-2(RNAi)	**15**	19.3 ± 0.9	26	60/65	-21	0	0.0004	0.7268	<0.0001$	1d
wild type (N2) L4440(RNAi)	**20**	22.9 ± 0.3	24	64/75						
wild type (N2) *daf-2*(RNAi)	**20**	38.9 ± 0.7	45	82/91	+69	+133	<0.0001	<0.0001	<0.0001$	
skn-1(zu135) L4440(RNAi)	**20**	16.7 ± 0.2	17	93/104	-27		<0.0001		<0.0001$	
skn-1(zu135) daf-2(RNAi)	**20**	16.6 ± 0.2	17	101/112	-28	-0.6	<0.0001	0.6062	<0.0001$	
daf-2(e1368) L4440(RNAi)	**20**	33.4 ± 0.9	38	51/63	+46	+100	<0.0001	<0.0001		
daf-2(e1368) daf-2(RNAi)	**20**	54.8 ± 0.9	64	62/73	+139	+228	<0.0001	<0.0001	<0.0001$	
daf-2(e1368); skn-1(zu135) L4440(RNAi)	**20**	21.6 ± 0.6	26	112/124	-6	+29	0.7408	<0.0001	<0.0001$	
daf-2(e1368); skn-1(zu135) daf-2(RNAi)	**20**	25.9 ± 1.2	36	81/94	+13	+55	0.0678	<0.0001	<0.0001$	
Trial of crude dauer pheromone at 25°C										
wild type (N2) L4440(RNAi) control	**25**	12.9 ± 0.4	14	30/30						
wild type (N2) L4440(RNAi) crude dauer pheromone	**25**	17.7 ± 0.8	21	30/30	+37	+50	<0.0001	<0.0001		
wild type (N2) *daf-2*(RNAi) control	**25**	21.1 ± 0.7	24	30/30	+64	+79	<0.0001	<0.0001		
wild type (N2) *daf-2*(RNAi) crude dauer pheromone	**25**	25.6 ± 0.9	28	29/30	+98	+117	<0.0001	<0.0001	<0.0001	
skn-1(zu135) L4440(RNAi) control	**25**	11.8 ± 0.3	13	29/30	-9		0.0235		<0.0001	1e
skn-1(zu135) L4440(RNAi) crude dauer pheromone	**25**	16.2 ± 0.6	19	29/30	+26	+37	<0.0001	<0.0001	<0.0001	1e
skn-1(zu135) daf-2(RNAi) control	**25**	12.2 ± 0.4	13	29/30	-5	+3	0.2311	0.4030	<0.0001	1e
skn-1(zu135) daf-2(RNAi) crude dauer pheromone	**25**	16.1 ± 0.6	21	29/30	+25	+36	<0.0001	<0.0001	<0.0001	1e

Lifespans were measured from the L4 stage, and animals that left the plates, buried into the agar, bagged, or exploded were censored. Analyses performed in parallel are grouped. L4440 empty vector was used as the RNAi control. Each *skn-1* mutant analysed is a strong loss-of-function and possible null. The class 2 alleles *daf-2*(*e1370*) and *daf-2*(*m596*) have comparably extended lifespans at 20 °C and 15 °C (Supplementary Table 2). The *daf-2*(*e1370*);*skn-1* double mutants lived 55% longer at 20 °C than at 15 °C (Supplementary Table 2), because *skn-1*-independent dauer-related processes increase their lifespan at the higher temperature (see text). This finding is striking given that *C. elegans* generally lives longer at lower temperatures[1] (Supplementary Table 2). Previous analyses of these transgenically rescued *daf-16* strains showed that DAF-16 expression specifically in neurons rescues the dauer but not longevity phenotypes of *daf-2*(*e1370*), whereas intestine-specific DAF-16 rescue allows lifespan extension but not dauer entry[15]. *N*, number of animals observed. *A merger of three trials shown in Supplementary Table 2. ED indicates data shown in an Extended Data Figure. *P* values were determined by log-rank. Additional experiments are shown in Supplementary Table 2.

Extended Data Table 2 | Collagen genes are upregulated by diverse interventions that increase lifespan

Experimental condition	Total # of genes upregulated	Reference	Enrichment score rank of collagens	# of collagens upregulated	# of collagens shared with *daf-2; skn-1* upregulated collagens	shared collagens tested in lifespan assays (in this study)
COLLAGENS UPREGULATED BY DRUG TREATMENTS THAT INCREASE *C. ELEGANS* LIFESPAN						
Resveratrol treatment in young wild-type adults	116	78	2	**8**	0	
Resveratrol treatment in young *daf-16(-)* adults	1027	78	1	**85**	28	*col-12, col-13, col-65, col-97, col-120, col-127, col-133,*
Humic acid treatment in 11 days old wild-type adults	740	79	1	**27**	5	*col-13, col-167, col-133*
Tannic acid treatment in young wild-type adults	2842	80	1	**74**	33	*col-10, col-12, col-13, col-65, col-97, col-133, col-141, col-144, col-167, col-180*
Quercetin treatment in young wild-type adults	1562	80	1	**67**	18	*col-12, col-13, col-97, col-133*
MAHMA (nitric oxide donor) in wild-type L4 worms	65	81	1	**8**	1	*col-97*
MAHMA (nitric oxide donor) in *hsf-1(sy441)* L4 worms	99	81	1	**21**	1	*col-97*
Rotenone treatment in young wild-type adults	2380	82	1	**64**	27	*col-10, col-65, col-97, col-133, col-141*
COLLAGENS UPREGULATED IN GENETIC BACKGROUNDS THAT INCREASE *C. ELEGANS* LIFESPAN						
Mixed-stage ***wdr-23****(tm1817)* mutants compared to wild type	2285	83	7	**41**	15	*col-10, col-144, col-167*
Young ***age-1****(mg44)* adults compared to wild type	791	84	1	**54**	9	*col-141*
daf-2(e1370) at day 5 of adulthood vs wild type*	869	85	1	**57**	19	*col-10, col-12, col-65, col-89, col-97, col-144, col-167*
daf-2*(m41)* at day 10 vs. wild type at day 6 of adulthood at 25.5°C	48	86	2	**17**	1	*col-141*
DAF-16-dependent genes expressed in ***daf-2(e1370)*** in day 1 adults at 20°C**	1078	74	1	**43**	16	*col-141*
TGFβ-dependent in day 1 adults	2181	74	1	**90**	30	*col-13, col-65, col-127, col-141, col-144, col-167, col-180*
AMPK and downstream signaling (shared transcriptional output of loss of *crh-1* (CREB) / loss of *tax-6* (calrectulin) / AAK-2 (AMPK) overexpression) in L4 larvae	549	87	1	**31**	17	*col-12, col-13, col-127, col-133, col-141, col-167, col-180*
ash-2 RNAi in animals that lacked a germline in day 8 adults	592	88	1	**21**	4	*col-12, col-133*
Young ***isp-1*** mutant adults compared to wild type	709	89	3	**15**	2	
cyc-1 RNAi in young adults	2459	89	1	**51**	18	*col-12, col-13, col-97, col-120, col-141, col-144*
2 day old ***rsks-1****(ok1255)* adults	155	90	1	**13**	3	*col-133*
Young ***ctbp-1****(ok498)* adults	213	91	1	**30**	16	*col-65, col-97, col-120, col-144, col-180*

Collagens were overrepresented in each *C. elegans* longevity-associated gene set we examined[74,78–91]. Gene Ontology (GO) enrichment clusters were identified by DAVID, using high-stringency classification. Enrichment scores were ranked from highest (1) to lowest (>10). Additional information is provided in Supplementary Table 10, including *P* values that were determined by DAVID using Fisher's exact test[53,92,93]. *Temperature not specified. **A comparison of *daf-2*(*e1370*) versus *daf-16*(*mu86*); *daf-2*(*e1370*).

Extended Data Table 3 | Suppression of lifespan extension by adulthood collagen gene knockdown

Strain / **RNAi**	Mean lifespan ± S.E.M. [Days]	75th percentile [Days]	N dead/ Initial N	% mean lifespan change to control	P-value (log-rank) vs. control	Figure
Trial of collagen genes from SKN-1-upregulated *daf-2(-)* set at 15°C						
rrf-3(pk1426) RNAi **L4440 (control)**	27.2 ± 0.4	29	55/64			ED Fig. 4c
rrf-3(pk1426) RNAi ***col-10***	28.2 ± 0.6	29	70/77	+4	0.0253	ED Fig. 4c
rrf-3(pk1426) RNAi ***col-65***	26.4 ± 0.5	29	57/70	-3	0.4555	ED Fig. 4c
rrf-3(pk1426) RNAi ***col-120***	25.7 ± 0.7	29	46/56	-6	0.2542	ED Fig. 4c
rrf-3(pk1426) RNAi ***col-127***	26.7 ± 0.5	29	48/61	-2	0.4797	ED Fig. 4c
rrf-3(pk1426) RNAi ***col-133***	27.2 ± 0.7	31	58/69	0	0.4623	ED Fig. 4c
rrf-3(pk1426) RNAi ***col-141***	25.3 ± 0.8	29	29/43	-7	0.0668	ED Fig. 4c
rrf-3(pk1426) RNAi ***col-167***	24.8 ± 0.7	29	47/64	-9	0.1128	ED Fig. 4c
rrf-3(pk1426) RNAi ***col-180***	27.3 ± 0.7	29	47/61	0	0.4668	ED Fig. 4c
P-value and % mean lifespan change are relative to *rrf-3(pk1426)* RNAi L4440						
daf-2(e1370); rrf-3(pk1426) RNAi **L4440**	37.3 ± 1.1	43	51/62			Fig. 3a
daf-2(e1370); rrf-3(pk1426) RNAi ***col-10***	28.6 ± 0.6	31	69/80	-23	<0.0001	Fig. 3a
daf-2(e1370); rrf-3(pk1426) RNAi ***col-65***	30.4 ± 0.8	36	71/86	-18	<0.0001	Fig. 3a
daf-2(e1370); rrf-3(pk1426) RNAi ***col-120***	26.6 ± 0.6	29	53/66	-29	<0.0001	Fig. 3a
daf-2(e1370); rrf-3(pk1426) RNAi ***col-127***	29.5 ± 0.7	33	56/62	-21	<0.0001	Fig. 3a
daf-2(e1370); rrf-3(pk1426) RNAi ***col-133***	28.2 ± 0.5	31	66/75	-24	<0.0001	Fig. 3a
daf-2(e1370); rrf-3(pk1426) RNAi ***col-141***	29.1 ± 0.6	31	58/71	-22	<0.0001	Fig. 3a
daf-2(e1370); rrf-3(pk1426) RNAi ***col-167***	28.3 ± 0.6	29	50/59	-24	<0.0001	Fig. 3a
daf-2(e1370); rrf-3(pk1426) RNAi ***col-180***	30.7 ± 0.7	33	53/65	-18	<0.0001	Fig. 3a
P-value and % mean lifespan change are relative to *daf-2(e1370); rrf-3(pk1426)* RNAi L4440						
eat-2(ad1116); rrf-3(pk1426) RNAi **L4440**	42.1 ± 1.0	47	75/81			Fig. 3c
eat-2(ad1116); rrf-3(pk1426) RNAi ***col-10***	38.9 ± 0.9	45	97/102	-7	0.0356	Fig. 3c
eat-2(ad1116); rrf-3(pk1426) RNAi ***col-65***	38.0 ± 1.0	43	75/79	-10	0.0038	Fig. 3c
eat-2(ad1116); rrf-3(pk1426) RNAi ***col-120***	37.1 ± 0.9	43	60/64	-12	<0.0001	Fig. 3c
eat-2(ad1116); rrf-3(pk1426) RNAi ***col-127***	37.4 ± 0.8	43	83/85	-11	<0.0001	Fig. 3c
eat-2(ad1116); rrf-3(pk1426) RNAi ***col-133***	37.3 ± 1.1	45	75/83	-11	0.0022	Fig. 3c
eat-2(ad1116); rrf-3(pk1426) RNAi ***col-141***	34.7 ± 1.2	43	49/54	-18	<0.0001	Fig. 3c
eat-2(ad1116); rrf-3(pk1426) RNAi ***col-167***	36.9 ± 1.0	43	62/69	-12	0.0002	Fig. 3c
eat-2(ad1116); rrf-3(pk1426) RNAi ***col-180***	34.5 ± 0.9	37	55/58	-18	<0.0001	Fig. 3c
P-value and % mean lifespan change are relative to *eat-2(ad1116); rrf-3(pk1426)* RNAi L4440						
Trial of collagen genes from SKN-1-upregulated *daf-2(-)* set at 20°C						
wild type (N2) RNAi **L4440 (control)**	24.6 ± 0.2	25	98/107			Fig. 3b
daf-2(e1370); rrf-3(pk1426) RNAi **L4440**	38.2 ± 0.8	44	86/93			Fig. 3b
daf-2(e1370); rrf-3(pk1426) RNAi ***col-10***	37.4 ± 0.7	42	104/110	-2	0.4014	Fig. 3b
daf-2(e1370); rrf-3(pk1426) RNAi ***col-13***	35.2 ± 0.7	42	92/99	-7	0.0043	Fig. 3b
daf-2(e1370); rrf-3(pk1426) RNAi ***col-120***	38.8 ± 0.6	44	103/110	+2	0.9386	Fig. 3b
P-value and % mean lifespan change are relative to *daf-2(e1370); rrf-3(pk1426)* RNAi L4440						
Trial of collagen genes from the SKN-1-upregulated *daf-2(-)* set at 20°C						
wild type (N2) RNAi **L4440 (control)** 0.2% DMSO	25.7 ± 0.3	29	96/103			Fig. 3d
wild type (N2) RNAi ***col-10*** 0.2% DMSO	25.7 ± 0.4	28	92/100	0	0.1545	
wild type (N2) RNAi ***col-13*** 0.2% DMSO	26.4 ± 0.3	27	102/111	+3	0.0811	
wild type (N2) RNAi ***col-120*** 0.2% DMSO	26.6 ± 0.2	29	98/104	+4	0.0147	
P-value and % mean lifespan change are relative to wild type (N2) RNAi L4440 (control) 0.2% DMSO						
wild type (N2) RNAi **L4440 (control)** 0.2% DMSO 100μM Rapamycin	30.7 ± 0.3	32	107/118			Fig. 3d
wild type (N2) RNAi ***col-10*** 0.2% DMSO 100μM Rapamycin	28.6 ± 0.5	32	83/93	-7	0.0083	Fig. 3d
wild type (N2) RNAi ***col-13*** 0.2% DMSO 100μM Rapamycin	28.0 ± 0.5	32	83/90	-8	0.0003	Fig. 3d
wild type (N2) RNAi ***col-120*** 0.2% DMSO 100μM Rapamycin	26.8 ± 0.5	29	77/87	-13	<0.0001	Fig. 3d
P-value and % mean lifespan change are relative to wild type (N2) RNAi L4440 (control) 0.2% DMSO 100μM Rapamycin						
Trial of collagen genes from SKN-1-upregulated *daf-2(-)* set at 20°C						
wild type (N2) RNAi **L4440 (control)**	23.8 ± 0.3	26	126/142			Fig. 3e
wild type (N2) RNAi ***col-10***	23.7 ± 0.3	26	94/108	0	0.9994	
wild type (N2) RNAi ***col-13***	23.3 ± 0.3	26	97/112	-2	0.2122	
wild type (N2) RNAi ***col-120***	23.1 ± 0.3	26	71/84	-3	0.1610	
P-value and % mean lifespan change are relative to wild type (N2) RNAi L4440 (control)						
glp-1(bn18) RNAi **L4440 (control)**	31.2 ± 0.6	34	53/64			Fig. 3e
glp-1(bn18) RNAi ***col-10***	25.8 ± 0.6	30	74/90	-17	<0.0001	Fig. 3e
glp-1(bn18) RNAi ***col-13***	27.8 ± 0.7	30	64/84	-11	0.0001	Fig. 3e
glp-1(bn18) RNAi ***col-120***	27.0 ± 0.6	30	75/98	-13	<0.0001	Fig. 3e
P-value and % mean lifespan change are relative to *glp-1(bn18)* RNAi L4440 (control)						

Lifespans were measured and presented as in Extended Data Table 1, and 1-day-old animals were placed on RNAi plates. Additional related experiments are shown in Supplementary Table 13. The *col-10* and *col-12* genes share more than 99% protein sequence identity with *col-144* and *col-13*, respectively. In analysis of *glp-1*(*bn18*), both N2 and *glp-1* animals were upshifted from 15 to 25 °C from the mid-L1 stage until the first day of adulthood, then we analysed lifespan at 20 °C.

doi:10.1038/nature14119

Axitinib effectively inhibits BCR–ABL1(T315I) with a distinct binding conformation

Tea Pemovska[1], Eric Johnson[2], Mika Kontro[3], Gretchen A. Repasky[1], Jeffrey Chen[2]†, Peter Wells[2], Ciarán N. Cronin[2], Michele McTigue[2], Olli Kallioniemi[1], Kimmo Porkka[3]§, Brion W. Murray[2]§ & Krister Wennerberg[1]§

The *BCR-ABL1* fusion gene is a driver oncogene in chronic myeloid leukaemia and 30–50% of cases of adult acute lymphoblastic leukaemia[1]. Introduction of ABL1 kinase inhibitors (for example, imatinib) has markedly improved patient survival[2], but acquired drug resistance remains a challenge[3–5]. Point mutations in the ABL1 kinase domain weaken inhibitor binding[6] and represent the most common clinical resistance mechanism. The BCR–ABL1 kinase domain gatekeeper mutation Thr315Ile (T315I) confers resistance to all approved ABL1 inhibitors except ponatinib[7,8], which has toxicity limitations. Here we combine comprehensive drug sensitivity and resistance profiling of patient cells *ex vivo* with structural analysis to establish the VEGFR tyrosine kinase inhibitor axitinib as a selective and effective inhibitor for T315I-mutant BCR–ABL1-driven leukaemia. Axitinib potently inhibited BCR–ABL1(T315I), at both biochemical and cellular levels, by binding to the active form of ABL1(T315I) in a mutation-selective binding mode. These findings suggest that the T315I mutation shifts the conformational equilibrium of the kinase in favour of an active (DFG-in) A-loop conformation, which has more optimal binding interactions with axitinib. Treatment of a T315I chronic myeloid leukaemia patient with axitinib resulted in a rapid reduction of T315I-positive cells from bone marrow. Taken together, our findings demonstrate an unexpected opportunity to repurpose axitinib, an anti-angiogenic drug approved for renal cancer, as an inhibitor for ABL1 gatekeeper mutant drug-resistant leukaemia patients. This study shows that wild-type proteins do not always sample the conformations available to disease-relevant mutant proteins and that comprehensive drug testing of patient-derived cells can identify unpredictable, clinically significant drug-repositioning opportunities.

The only clinically available inhibitor that has shown efficacy against BCR–ABL1(T315I)-driven disease is the broad-spectrum kinase inhibitor ponatinib. In a recent phase III clinical trial with ponatinib, frequent severely adverse vascular effects were observed leading to termination of the trial as well as temporary withdrawal from the market[9]. Moreover, compound mutations in BCR–ABL1(T315I) have been reported to cause resistance to ponatinib[10,11]. Therefore, there is a significant, unmet need for safe and effective therapies for BCR–ABL1(T315I)-driven leukaemia.

We performed phenotypic drug sensitivity and resistance testing (DSRT)[12,13] of primary cells from chronic myeloid leukaemia (CML) and Philadelphia-chromosome-positive (Ph^+) acute lymphoblastic leukaemia (ALL) patients using a collection of 252 approved and investigational oncology compounds (Supplementary Table 1). DSRT data from a Ph^+ B-cell ALL (B-ALL) patient (FHRB.1278) carrying the T315I mutation showed that the patient cells were insensitive to imatinib, dasatinib and nilotinib, and sensitive to ponatinib (Fig. 1a, b and Supplementary Table 2). Anticipated cancer-selective responses were observed to drugs targeting key BCR–ABL1 downstream effector signals such as PI(3)K/MTOR, MEK and BCL2 (Supplementary Table 2).

Unexpectedly, a strong selective response was detected with the vascular endothelial growth factor (VEGF) receptor tyrosine kinase inhibitor axitinib. Highly selective for VEGFR[14], axitinib is approved as an anti-angiogenic agent for treating renal cell carcinoma[15]. No other VEGFR inhibitors in our drug collection showed activity in these cells, suggesting that the effect of axitinib was due to inhibition of a different target. Since axitinib had not been previously investigated in cell-based systems for potency and selectivity towards ABL1, we explored whether the observed *ex vivo* effects were due to direct targeting of the BCR–ABL1(T315I) kinase and characterized the biochemical interactions. Surprisingly, axitinib more potently inhibited the kinase activity of ABL1(T315I) (inhibition constant (K_i) = 100 pM), in line with VEGFR2 potency (K_i = 20 pM)[14,16], in comparison to wild-type ABL1 (K_i = 3,800 pM).

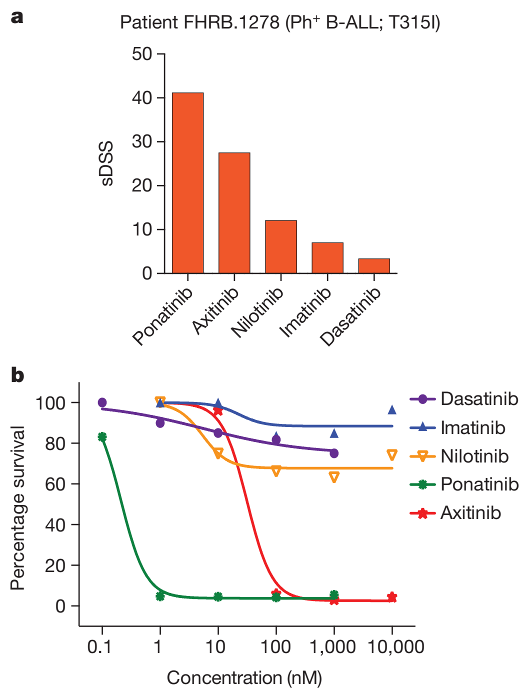


Figure 1 | *Ex vivo* drug sensitivity and resistance testing (DSRT) of primary leukaemic cells establishes axitinib as a selective BCR-ABL1(T315I) inhibitor. **a**, DSRT of Ph^+ B-ALL patient cells (FHRB. 1278) identifies axitinib as a potent leukaemia-selective drug. Data are shown in a waterfall plot as selective drug sensitivity score (sDSS)[12,13] in which patient cell drug-response data are compared with responses of healthy donor bone marrow mononuclear cell control samples ($n = 7$). The sDSS values of other BCR–ABL1 inhibitors are provided for comparison. **b**, Dose-response data of axitinib and ponatinib, imatinib, dasatinib and nilotinib as percentage survival.

[1]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, 00290 Helsinki, Finland. [2]La Jolla Laboratories, Pfizer Worldwide Research & Development, San Diego, California 92121, USA. [3]Hematology Research Unit Helsinki, University of Helsinki, and Helsinki University Hospital Comprehensive Cancer Center, Department of Hematology, 00290 Helsinki, Finland. †Present address: Wellspring Biosciences LLC, La Jolla, California 92037, USA.
§These authors jointly supervised this work.

Crystal structures of axitinib bound to wild-type and T315I ABL1 were determined (Extended Data Table 1) to understand binding interactions and selectivity. The structures revealed a striking difference in the activation loop (A-loop) conformations. While the wild-type ABL1 bound axitinib in an inactive (DFG-out conformation), as has also been shown for the axitinib–VEGFR complex, the axitinib–ABL1(T315I) complex had an active (DFG-in) A-loop conformation (Fig. 2a, b).

As the T315I gatekeeper mutation is known to stabilize the active kinase conformation[17,18], the difference in A-loop conformations between the wild type and the T315I mutant probably reflects the altered protein dynamics. The higher potency against the mutant implies that axitinib binds better to the active ABL kinase conformation. This was corroborated with biochemical studies where axitinib more potently inhibited autophosphorylated ($K_i = 149$ pM) rather than non-phosphorylated ABL1(T315I) ($K_i = 421$ pM).

Moreover, axitinib filled different binding space than other ABL1 inhibitors (ponatinib, imatinib, dasatinib, nilotinib, bosutinib), as it did not extend as far towards the gatekeeper residue and α-helix C (Fig. 2c). Also, unlike other ABL1 inhibitors that rely on forming a hydrogen bond with the T315 residue and thus lose potency with an isoleucine substitution[19], axitinib was not positioned to form a hydrogen bond with T315 and therefore did not clash unfavourably with I315. These findings reveal the structural underpinnings of potent, ABL1(T315I)-selective inhibition, which can be used in the rational design of a new generation of ABL1-directed drugs.

Unexpectedly, axitinib adopts a significantly different binding conformation in the active site of ABL1 relative to VEGFR2[14], specifically to the P-loop and A-loop conformation of the kinases (Fig. 2d and Extended Data Fig. 1). A large rotational difference of the axitinib sulfur–indazole bond placed the phenyl amide group in different binding pockets.

Next, to test whether axitinib specifically blocks ABL1(T315I) kinase activity in cells we used murine pro-B Ba/F3 cells stably expressing either wild-type or mutant BCR–ABL1 to examine ABL1 autophosphorylation and BCR–ABL1-dependent cell proliferation. Similar to ponatinib, axitinib potently reduced autophosphorylation of ABL1(T315I) and potently blocked proliferation of Ba/F3 cells expressing BCR–ABL1(T315I) in a dose-dependent manner (Fig. 3a, b). Strikingly, axitinib inhibited T315I-mediated autophosphorylation and Ba/F3 growth with ~10-fold higher potency compared to wild-type ABL1 (Fig. 3c, d). Finally, in a profile against a large set of clinically relevant BCR–ABL1 drug-resistance mutations expressed in Ba/F3 cells, axitinib showed high selective inhibitory activity towards gatekeeper mutations (Extended Data Fig. 2). Taken together, our data demonstrate that axitinib selectively and effectively targets gatekeeper-mutant BCR–ABL1 on a biochemical, structural and cellular level.

To assess the efficacy of axitinib in patient samples further, we performed DSRT on mononuclear cells isolated from the bone marrow of two CML patients negative for the T315I mutation as well as three additional patients carrying the T315I mutation (patient characteristics described in Extended Data Table 2). When compared with a panel of 32 primary acute myeloid leukaemia samples and seven healthy bone marrow controls, samples derived from patients with BCR–ABL1(T315I)-driven leukaemia showed a selective response to axitinib, confirming our observations with the Ph^+ B-ALL patient (FHRB.1278) (Fig. 4a and Extended Data Table 3). In accordance with the biochemical and Ba/F3 cell viability results, axitinib selectively and effectively inhibited the viability of three of four CML and Ph^+ ALL primary cell samples harbouring the T315I mutation with median IC_{50} values 20-fold lower (Fig. 4b and Supplementary Table 3). The axitinib-insensitive patient sample was also insensitive to ponatinib, suggesting that the patient had a BCR–ABL1-independent disease (Fig. 4c).

We then focused on one of the tested CML patients (FHRB.1408) for whom all approved treatment options had been exhausted and *ex vivo* cell viability analysis had shown a solid response to axitinib (Extended Data Fig. 3a, b). To determine whether axitinib could block BCR–ABL1(T315I)-mediated signalling in the patient's tumour cells, we assessed phosphorylation of the BCR–ABL1 substrate CRKL. *Ex vivo* axitinib treatment caused a dose-dependent reduction of CRKL phosphorylation, indicating that axitinib effectively inhibited BCR–ABL1 (T315I) signalling in the primary leukaemic cells (Fig. 4d). Based on the potent on-target *ex vivo* responses, the patient was treated with an approved therapeutic dose of axitinib (5 mg twice daily) for two weeks.

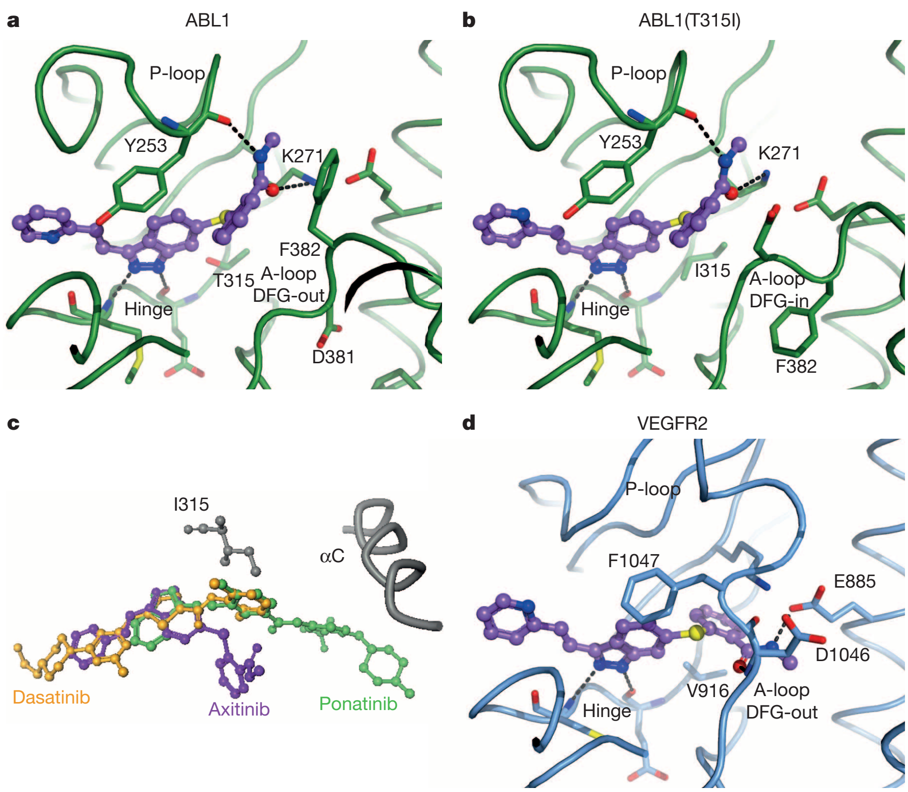


Figure 2 | Co-crystal structures of axitinib bound to ABL1, ABL1(T315I) and VEGFR2 demonstrating different binding conformations. Axitinib forms four hydrogen bonds (dashed lines) with ABL1, two between the ABL1 kinase hinge segment and the axitinib indazole ring and two between the K271 and Y253 residues and the axitinib amide. **a**, Crystal structure of axitinib (purple) in complex with ABL1 where the A-loop is in an inactive (DFG-out) conformation. **b**, Axitinib–ABL1(T315I) complex where the A-loop is in an active DFG-in conformation. **c**, Overlay of the ABL1-bound conformation of axitinib, dasatinib (PDB ID: 2GQG) and ponatinib (PDB ID: 3IK3). αC, α-helix C. **d**, Axitinib bound to VEGFR2 with a DFG-out A-loop.

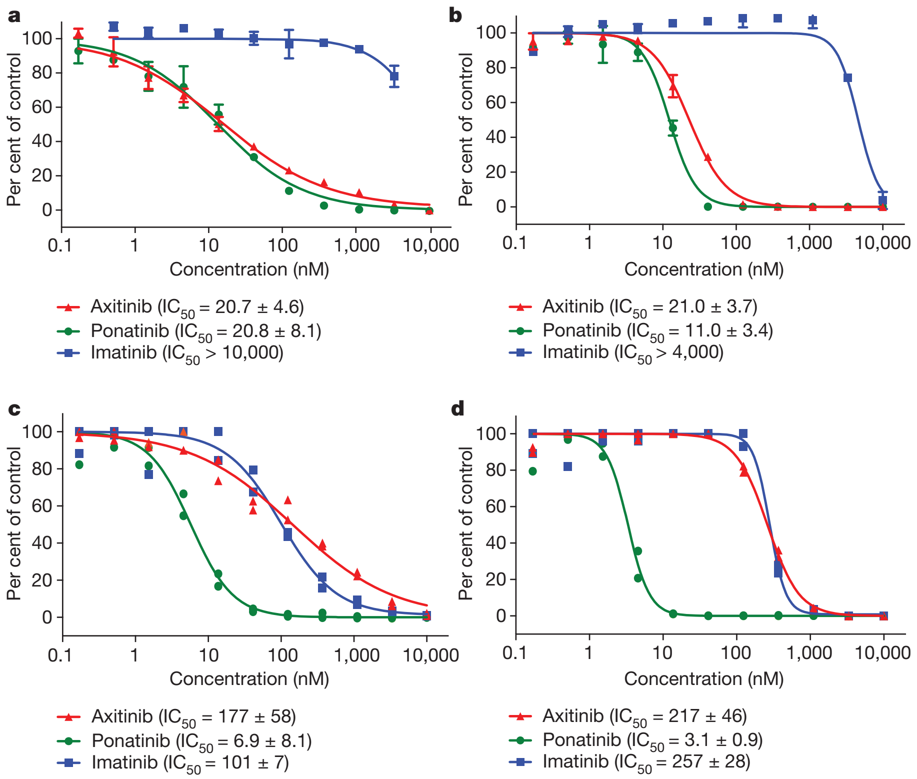


Figure 3 | Axitinib suppresses BCR–ABL1(T315I) autophosphorylation and proliferation of Ba/F3 cells expressing BCR–ABL1(T315I). **a**, Imatinib-, axitinib- and ponatinib-mediated inhibition of autophosphorylation of ABL1(T315I) kinase in engineered Ba/F3 cell lines measured with a phospho-ABL1 enzyme-linked immunosorbent assay (ELISA). **b**, Oncogene-dependent proliferation of Ba/F3 cell lines expressing BCR–ABL1(T315I) following 96 h treatment with increasing concentrations of imatinib, axitinib or ponatinib quantified with a resazurin cell viability assay. **c**, **d**, Imatinib-, axitinib- and ponatinib-mediated inhibition of autophosphorylation of ABL1 kinase and proliferation of Ba/F3 cells expressing BCR–ABL1. Half-maximal inhibitory concentrations (IC_{50}) detected from multiple repeated experiments are shown below each graph. **a**, **b**, $n = 3$; **c**, **d**, $n = 2$. Error bars show standard deviations.

The treatment resulted in rapid clearance of BCR–ABL1(T315I)-positive cells as determined by a reduction of the T315I transcript levels in bone marrow (Fig. 4e). This finding suggests that axitinib can produce specific and effective responses in patients with BCR–ABL1(T315I)-driven disease.

Although the majority of patients with chronic phase CML achieve a significant therapeutic benefit with clinically available BCR–ABL1 inhibitors, there is an unmet medical need for novel therapies for a subgroup of patients who develop resistance to treatment. Here we employ a wide array of approaches (crystallographic, biochemical, *ex vivo* cellular, and *in vivo*) to demonstrate that axitinib selectively targets BCR–ABL1(T315I) through a gatekeeper-mutant-selective mechanism. Although several drug candidates have been reported to target this gatekeeper mutation, including the approved drug ponatinib[20–23], their clinical utility has typically been limited by toxicity. Axitinib is a valuable addition to the BCR–ABL1 armamentarium because it is (1) selective towards the gatekeeper mutant, (2) binds to the T315I active site in a distinctive manner as compared to other ABL1 drugs, and (3) has a narrow target profile[14,23] (Extended Data Fig. 4), which could translate into fewer and less severe side effects. This narrow-profile gatekeeper-selective mechanism of action highlights axitinib as a new type of ABL1 kinase inhibitor. This finding could pave the way for a new approach for drug development towards even more potent and selective gatekeeper-mutant inhibitors targeting ABL1 as well as other relevant kinases such as EGFR and KIT and their mutation-specific conformations.

Since axitinib is already approved for patients with refractory renal cell carcinoma and has manageable side effects[24–26], our study provides a solid rationale for formal exploration of the clinical utility of axitinib in drug-resistant BCR–ABL1(T315I)-driven leukaemia in a fast-track mode, probably in combination with a conventional ABL1 inhibitor.

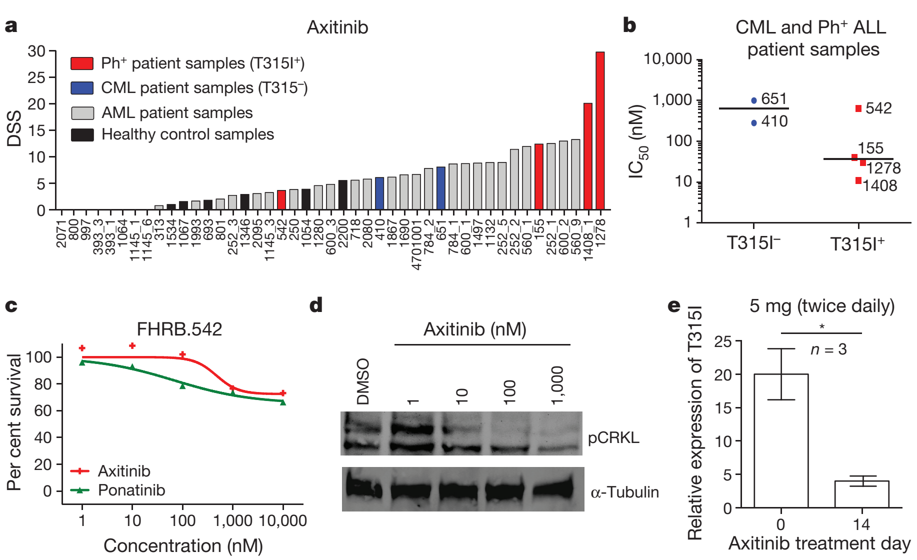


Figure 4 | Axitinib potently and selectively targets BCR–ABL1(T315I)-expressing patient cells. **a**, Distribution of *ex vivo* sensitivity (represented as DSS) to axitinib in primary BCR–ABL1(T315)-positive, BCR–ABL1(T315I)-negative CML, AML and healthy bone marrow control samples. **b**, Comparison of axitinib half maximal inhibitory concentrations (IC_{50}) in T315I negative versus positive CML and Ph^+ ALL patient samples. Lines indicate median. **c**, Dose response of axitinib and ponatinib in patient sample FHRB.542. **d**, Immunoblot for CRKL phosphorylation following exposure to axitinib in patient sample FHRB.1408. **e**, Clinical response of patient FHRB.1408 to axitinib. Real-time quantitative PCR measurement of the relative percentage of BCR–ABL1(T315I) transcripts in blood before and after 14 days of clinical treatment with axitinib. Error bars indicate 19.2% in-between run coefficient of variation; the P value ($*P < 0.05$) signifies the difference in BCR–ABL1(T315I) transcript levels at treatment days 0 and 14 (Unpaired t-test).

Axitinib may represent a unique opportunity for combinations with other ABL1 drugs because it does not have a similar adverse event profile or overlapping toxicities[27,28]. Furthermore, most current ABL1 drugs cause inactivation of the enzyme that metabolizes them (cytochrome P450, family 3, subfamily A, polypeptide 4) that would complicate the prediction of drug exposures in combinations. Axitinib and bosutinib do not have this issue and their combination may enable broad-spectrum therapy. It is also plausible that the adverse effects of axitinib (mainly VEGFR-driven) could be minimized by intermittent exposure, which may be sufficient for treating BCR–ABL1(T315I)-driven leukaemia as indicated with other ABL1 inhibitors[29,30].

In summary, our results on axitinib in T315I mutant CML provide a powerful example of how unbiased drug sensitivity testing of patient-derived cancer cells can lead to the discovery of an unexpected drug–target interaction with mechanistic, structural and clinical implications. Our study highlights the value of drug repositioning, that is, searching for novel indications for existing, emerging and abandoned drugs, such as in the NIH program 'Discovering New Therapeutic Uses for Existing Molecules'.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 July; accepted 26 November 2014.
Published online 9 February 2015.

1. Kurzrock, R., Kantarjian, H. M., Druker, B. J. & Talpaz, M. Philadelphia chromosome-positive leukemias: from basic mechanisms to molecular therapeutics. *Ann. Intern. Med.* **138,** 819–830 (2003).
2. Deininger, M., Buchdunger, E. & Druker, B. J. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood* **105,** 2640–2653 (2005).
3. Druker, B. J. *et al.* Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N. Engl. J. Med.* **355,** 2408–2417 (2006).
4. Hochhaus, A. *et al.* Dasatinib induces notable hematologic and cytogenetic responses in chronic-phase chronic myeloid leukemia after failure of imatinib therapy. *Blood* **109,** 2303–2309 (2007).
5. Hochhaus, A. *et al.* Six-year follow-up of patients receiving imatinib for the first-line treatment of chronic myeloid leukemia. *Leukemia* **23,** 1054–1061 (2009).
6. Branford, S. *et al.* High frequency of point mutations clustered within the adenosine triphosphate-binding region of BCR/ABL in patients with chronic myeloid leukemia or Ph-positive acute lymphoblastic leukemia who develop imatinib (STI571) resistance. *Blood* **99,** 3472–3475 (2002).
7. Bradeen, H. A. *et al.* Comparison of imatinib mesylate, dasatinib (BMS-354825), and nilotinib (AMN107) in an N-ethyl-N-nitrosourea (ENU)-based mutagenesis screen: high efficacy of drug combinations. *Blood* **108,** 2332–2338 (2006).
8. Redaelli, S. *et al.* Activity of bosutinib, dasatinib, and nilotinib against 18 imatinib-resistant BCR/ABL mutants. *J. Clin. Oncol.* **27,** 469–471 (2009).
9. Senior, M. FDA halts then allows sales of Ariad's leukemia medication. *Nature Biotechnol.* **32,** 9–11 (2014).
10. Gibbons, D. L. *et al.* Molecular dynamics reveal BCR-ABL1 polymutants as a unique mechanism of resistance to PAN-BCR-ABL1 kinase inhibitor therapy. *Proc. Natl Acad. Sci. USA* **111,** 3550–3555 (2014).
11. Zabriskie, M. S. *et al.* BCR-ABL1 compound mutations combining key kinase domain positions confer clinical resistance to ponatinib in Ph chromosome-positive leukemia. *Cancer Cell* **26,** 428–442 (2014).
12. Pemovska, T. *et al.* Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. *Cancer Discov.* **3,** 1416–1429 (2013).
13. Yadav, B. *et al.* Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. *Sci. Rep.* **4,** 5193 (2014).
14. McTigue, M. *et al.* Molecular conformations, interactions, and properties associated with drug efficiency and clinical performance among VEGFR TK inhibitors. *Proc. Natl Acad. Sci. USA* **109,** 18281–18289 (2012).
15. Rini, B. I. *et al.* Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. *Lancet* **378,** 1931–1939 (2011).
16. Solowiej, J. *et al.* Characterizing the effects of the juxtamembrane domain on vascular endothelial growth factor receptor-2 enzymatic activity, autophosphorylation, and inhibition by axitinib. *Biochemistry* **48,** 7019–7031 (2009).
17. Azam, M., Seeliger, M. A., Gray, N. S., Kuriyan, J. & Daley, G. Q. Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nature Struct. Mol. Biol.* **15,** 1109–1118 (2008).
18. Dixit, A. & Verkhivker, G. M. Hierarchical modeling of activation mechanisms in the ABL and EGFR kinase domains: thermodynamic and mechanistic catalysts of kinase activation by cancer mutations. *PLOS Comput. Biol.* **5,** e1000487 (2009).
19. Zhou, T. *et al.* Structural mechanism of the pan-BCR-ABL inhibitor ponatinib (AP24534): lessons for overcoming kinase inhibitor resistance. *Chem. Biol. Drug Des.* **77,** 1–11 (2011).
20. O'Hare, T. *et al.* AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. *Cancer Cell* **16,** 401–412 (2009).
21. Gontarewicz, A. *et al.* Simultaneous targeting of Aurora kinases and Bcr-Abl kinase by the small molecule inhibitor PHA-739358 is effective against imatinib-resistant *BCR-ABL* mutations including T315I. *Blood* **111,** 4355–4364 (2008).
22. Chan, W. W. *et al.* Conformational control inhibition of the BCR-ABL1 tyrosine kinase, including the gatekeeper T315I mutant, by the switch-control inhibitor DCC-2036. *Cancer Cell* **19,** 556–568 (2011).
23. Davis, M. I. *et al.* Comprehensive analysis of kinase inhibitor selectivity. *Nature Biotechnol.* **29,** 1046–1051 (2011).
24. Gross-Goupil, M., Francois, L., Quivy, A. & Ravaud, A. Axitinib: a review of its safety and efficacy in the treatment of adults with advanced renal cell carcinoma. *Clin. Med. Insights Oncol.* **7,** 269–277 (2013).
25. Bracarda, S. *et al.* Axitinib safety in metastatic renal cell carcinoma: suggestions for daily clinical practice based on case studies. *Expert Opin. Drug Saf.* **13,** 497–510 (2014).
26. Verzoni, E. *et al.* Targeted treatments in advanced renal cell carcinoma: focus on axitinib. *Pharmgenomics. Pers. Med.* **7,** 107–116 (2014).
27. Josephs, D. H., Fisher, D. S., Spicer, J. & Flanagan, R. J. Clinical pharmacokinetics of tyrosine kinase inhibitors: implications for therapeutic drug monitoring. *Ther. Drug Monit.* **35,** 562–587 (2013).
28. Chen, Y. *et al.* Clinical pharmacology of axitinib. *Clin. Pharmacokinet.* **52,** 713–725 (2013).
29. Shah, N. P. *et al.* Transient potent BCR-ABL inhibition is sufficient to commit chronic myeloid leukemia cells irreversibly to apoptosis. *Cancer Cell* **14,** 485–493 (2008).
30. O'Hare, T. *et al.* Threshold levels of ABL tyrosine kinase inhibitors retained in chronic myeloid leukemia cells determine their commitment to apoptosis. *Cancer Res.* **73,** 3356–3370 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank the patients and their families for participating in this study and donating their samples for research. We acknowledge the High Throughput Biomedicine Unit at the Institute for Molecular Medicine Finland (FIMM) for technical assistance, the laboratory of C. Gambacorti-Passerini (University of Milano-Bicocca) for engineered Ba/F3 cell mutant panel profiling data, and T. Lundán (Department of Clinical Chemistry and TYKSLAB, Turku University Central Hospital, University of Turku) for the BCR–ABL1(T315I) transcript-level quantification. This work was supported by the Jane and Aatos Erkko Foundation (to K.W.), Academy of Finland (K.W. and O.K.), Finnish Cancer Societies (O.K. and K.P.), Sigrid Juselius Foundation (O.K.), Instrumentarium Foundation (M.K.), and FinPharma Doctoral Program-Drug Discovery section (T.P.).

Author Contributions T.P., K.P., B.W.M. and K.W. conceived the study, designed experiments and wrote the manuscript. T.P. performed the DSRT and *ex vivo* assays on patient samples and associated data analysis. E.J., C.C., M.M. and B.W.M. designed, performed and interpreted the crystallography experiments. M.K. and K.P. coordinated the sampling of patient material, collection of associated clinical data, and clinical translation. G.A.R. and O.K. contributed to study design and manuscript writing. J.C., P.W. and B.W.M. coordinated and performed the *in vitro* biochemical and cellular experiments. K.P., B.W.M. and K.W. supervised the experimental and clinical analysis. All authors discussed the results, commented and edited the manuscript.

Author Information X-ray crystallographic coordinates and structure factor files for axitinib–ABL1(T315I) and axitinib–wild-type-ABL1 complexes have been deposited in the Protein Data Bank under accession numbers 4TWP and 4WA9, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.W. (krister.wennerberg@fimm.fi) or B.W.M. (brion.murray@pfizer.com).

METHODS

Study patients and material. Primary patient material was obtained after written informed consent approved by the Helsinki University Central Hospital Institutional Review Board (No. 239/13/00/2010, 303/13/03/01/2011). The consent also included the possibility to utilize DSRT data to guide therapies with approved agents in an off-label manner in accordance with Finnish legislation. Mononuclear cells from bone marrow aspirates of CML and Ph^+ ALL patients were isolated using Ficoll centrifugation (Ficoll-Paque PREMIUM; GE Healthcare) and maintained in Mononuclear Cell Medium (MCM; PromoCell) supplemented with 0.5 $\mu g\ ml^{-1}$ gentamicin and 2.5 $\mu g\ ml^{-1}$ amphotericin B. Patient characteristics are summarized in Extended Data Table 2.

Inhibitors and DSRT. The oncology compound collection used in this study consisted of 125 FDA/EMA approved anti-cancer drugs, along with 127 investigational and preclinical compounds covering a wide spectrum of molecular targets (Supplementary Table 1). All of the compounds were purchased from commercial chemical vendors and dissolved in either 100% dimethyl sulfoxide (DMSO) or water. The DSRT was performed as previously described[12]. Specifically, each compound was tested in five different concentrations covering a 10,000-fold concentration range and preprinted on 384 microtitre tissue culture treated plates (Corning) with an acoustic liquid handling device (Echo 550, Labcyte Inc.). Five microlitres of medium was added to each well for compound dissolution and the plates were gently shaken for 30 min. A single-cell suspension of freshly isolated mononuclear cells (20 µl per well; 10,000 cells per well) was then transferred to every well using Multi-Drop Combi peristaltic dispenser (Thermo Scientific). The 384-well plates were then incubated for 72 h at 37 °C and 5% CO_2 and following the incubation period cell viability was measured using CellTiter-Glo reagent (Promega) according to the manufacturer's instructions with Molecular Devices Paradigm plate reader. Cell viability luminescence data was normalized to DMSO-only wells (negative control) and 100 µM benzethonium-chloride-containing wells (positive control). Dose-response curves were generated in Dotmatics Browser/Studies software (Dotmatics Ltd) on the basis of four-parameter logistics fit function (minimum and maximum inhibition, slope and half maximal inhibitory concentration (IC_{50})). The DSRT data were evaluated with a custom-developed drug sensitivity score (DSS)[12,13]. Selective drug responses in the CML and Ph^+ ALL patient cells were evaluated in comparison with the average control cell drug sensitivity profile, sDSS.

Kinase assays. Wild-type ABL1 and ABL1(T315I) enzyme inhibition was measured using a microfluidic mobility shift assay. The reactions were conducted in 50 µl volumes in 96-well plates and contained GST-tagged human-recombinant ABL1(T315I) kinase intracellular domain (1 nM), 3 µM phosphoacceptor peptide, 5′ FAM-EAIYAAPFAKKK-OH (CPC Scientific, also known as ProfilerPro Peptide 2, Caliper Life Sciences), test compound (11-dose threefold serial dilutions, 2% DMSO final) or DMSO only, 1 mM dithiothreitol (DTT), 0.002% Tween-20 and 5 mM $MgCl_2$ in 25 mM HEPES, pH 7.1. The reactions were initiated by addition of ATP (5 µM final concentration, which is approximately equal to the Michaelis constant (K_m)) following a 20 min pre-incubation, incubated for 1.5 h at room temperature and stopped by the addition of 0.1 M EDTA, pH 8. Extent of reaction (~15–20% conversion with no inhibitor) was determined after electrophoretic separation of the fluorescently labelled peptide substrate and phosphorylated product on a LabChip EZ Reader II (Caliper Life Sciences). The K_i values were calculated by fitting the percentage conversion to the Morrison[31] equation for tight-binding competitive inhibition using a nonlinear regression method (GraphPad Prism), best fit enzyme concentration value and an experimentally measured ATP K_m (~4 µM).

Production of recombinant ABL1. The nucleotide sequence encoding residues 229–515 of human ABL1a (NM_005157.4) was obtained from GenScript sub-cloned into an insect cell expression transfer vector that appended the N-terminus with the tobacco etch virus (TEV)-cleavable polyhistidine purification tag sequence MASHHHHHHDYDGATTENLYFQ/GS, where TEV cleaves at the solidus, leaving the recombinant protein with a Gly–Ser extension at the N-terminus. Site-directed mutagenesis was used to generate the T315I mutation within this construct. TEV protease was produced in-house under license from NCI[32]. Recombinant baculoviruses were prepared by using the 'Bac-to-Bac' method (Invitrogen) and used to infect 10 l of Sf21 insect cells at 27 °C and at a multiplicity of infection (MOI) of 1. The ABL viruses were co-infected with a baculovirus expressing YopH tyrosine phosphatase at a MOI of 0.01, in order to generate the non-phosphorylated protein species. Infected cells were harvested 72 h post-infection and the PBS-washed cell pellets were stored at −80 °C before purification.

Frozen insect cell pellets containing recombinant ABL1(T315I) mutant kinase domain protein were resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, 5 mM ADP-NaOH pH 7.5, 0.25 mM TCEP, 2 µM leupeptin (Sigma Chemical), and one 'EDTA-free' protease inhibitor tablet (Roche) per 75 ml buffer), and the mixture stirred at 4 °C for 1 h followed by centrifugation at 5,000*g* for 1 h. The supernatant fraction was incubated with 5 ml of ProBond resin (Invitrogen) for approximately 3 h with mixing at 4 °C. Subsequently, the resin containing bound ABL1 was batch washed with 4 × 50 ml of wash buffer (50 mM Tris-HCl, pH 8.0, 400 mM NaCl, 20 mM imidazole-HCl, pH 8.0, 10 mM $MgCl_2$, 2.5 mM ADP-NaOH pH 7.5, 0.25 mM TCEP, 1 µM leupeptin) and the resin transferred to a disposable Econo column (Bio-Rad). The resin was washed further with 20 ml of wash buffer and the bound protein step-eluted by using four column volumes of elution buffer (50 mM Tris-HCl, pH 8.0, 400 mM NaCl, 250 mM imidazole-HCl, pH 8.0, 10 mM $MgCl_2$, 2.5 mM ADP-NaOH pH 7.5, 0.25 mM TCEP, 1 µM leupeptin). The eluted protein was treated with TEV protease during overnight dialysis against wash buffer. The dialysed material was passed through a fresh 10 ml column of ProBond resin previously equilibrated with the post-dialysis buffer and the flow-through fraction containing the detagged ABL1(T315I) was collected. The protein solution was dialysed against delivery buffer (25 mM HEPES-NaOH, pH 7.2, 250 mM NaCl, 5 mM $MgCl_2$, 2.5 mM ADP-NaOH, pH 7.5, 20% (v/v) glycerol, 0.25 mM TCEP). The wild-type ABL1 kinase domain was purified in a similar manner with the exception that $MgCl_2$ and ADP-NaOH pH 7.0 were omitted from all buffers. The purified proteins at 2–3 $mg\ ml^{-1}$ were flash frozen in liquid N_2 and stored at −80 °C. Protein measurements were determined by using the Coomassie Plus Protein Reagent (Pierce).

Enzyme preparation. Frozen wild-type ABL1 and ABL1(T315I) proteins were thawed and diluted to 10 µM (~0.3 $mg\ ml^{-1}$) with cold 25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM DTT. An equal volume of cold buffer containing 20 µM axitinib (freshly diluted from a 20 mM stock in DMSO) was mixed with the protein and placed on ice for 1 h. The protein mixture was concentrated to ~20 $mg\ ml^{-1}$ by using a Millipore centrifugal concentrator and the proteins flash frozen in liquid N_2 before crystallization trials.

Crystallization. All crystals were obtained by sitting drop vapour diffusion in SBS format MRC2 crystallization plates using a Mosquito liquid handler. Crystals of ABL(T315I) were obtained at 13 °C by mixing 196 nl protein–axitinib complex solution (17 $mg\ ml^{-1}$) with 211 nl of reservoir solution (15.0% (w/v) PEG 3350, 10 mM $MgCl_2$, 5 mM $NiCl_2$, 5.0% (v/v) glycerol, and 100 mM HEPES, pH 7). Wild-type ABL1 crystals grew from drops containing 150 nl protein (18 $mg\ ml^{-1}$) mixed with 190 nl of reservoir solution (0.1 M ammonium chloride, 20.0% (w/v) PEG 3350 and 5.0% (v/v) ethylene glycol). Prior to crystal harvest, the crystallization drop was covered with 4 µl of reservoir solution containing 20% glycerol (T315I) or 20% ethylene glycol (wild type) as cryo-protectant. Crystals were then harvested directly from the crystallization drop followed by immediate flash-freezing in liquid N_2.

Data collection and structure determination. X-ray data sets were collected at beamline 17-ID at the Advanced Photon Source synchrotron (Argonne National Laboratories) using a wavelength of 1 Å, and at a temperature of 100 K with $R_{merge} = 0.055$ for data in the range of 65.58–2.40 Å for ABL1(T315I), and $R_{merge} = 0.057$ for data in the range of 111.81–2.20 Å for wild-type ABL1 (Extended Data Table 1). Structure solution for ABL1(T315I) proceeded by molecular replacement in PHASER[33] using the ABL1 structure (PDB ID: 3IK3) as the starting model followed by iterative rounds of model building and refinement using Coot[34] and REFMAC5[35]. Final refinement and structure validation were performed in the PHENIX suite[36]. Structure solution, model building and refinement for wild-type ABL1 proceeded in an identical manner but using the refined ABL1(T315) atomic coordinates as a starting model for molecular replacement. Crystallographic *R* factors and stereochemistry statistics indicate high quality models for each refinement (Extended Data Table 1). For ABL1(T315I) and wild-type ABL1 structures respectively Ramachandran plots indicate 97.9% and 97.1% of residues are in the favoured, and 2.1% and 2.9% in the allowed, regions. Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank under accession numbers 4TWP and 4WA9 for ABL1(T315I) and wild-type ABL1 respectively.

BCR–ABL1 autophosphorylation ELISA. Murine pro-B Ba/F3 cell lines expressing human wild-type BCR–ABL1 and mutant BCR–ABL1(T315I) were obtained from Oregon Health and Science University and grown in RPMI-1640 medium supplemented with 10% FBS and 1% penicillin–streptomycin. For the ABL1 phospho-Tyr ELISA, cells were pipetted into a 50 ml tube, centrifuged at 180*g*, and the cell pellet was re-suspended in assay medium (RPMI-1640 with 0.1% FBS, 0.05% BSA w/v, and 1% penicillin–streptomycin). The cells were counted using an Innovatis Cedex Cell Counter, seeded into a 96-well flat-bottom plate in assay medium at 40,000 cells per well and incubated for 2 h at 37 °C, 5% CO_2 95% air. Test compounds were dissolved in 100% DMSO to make 10 mM stocks. The compounds were then diluted in 100% DMSO in a polypropylene 96-well plate using threefold serial dilution. Control wells contained 100% DMSO without test compound (uninhibited controls). The DMSO drug dilution plate was diluted 40-fold into assay medium to yield a 5× drug source plate for the assay. Twenty-five microlitres was transferred from the 5× source plate to the cell assay plate and the assay plate was incubated with test compounds for an additional 2 h at 37 °C, 5% CO_2 95% air. Following this incubation, the cells were centrifuged at 405*g* for 5 min and 80 µl of

the supernatant was removed. Cells were lysed by adding 100 μl per well of freshly prepared Cell Signaling Technology lysis buffer (#9803) supplemented with 1% SDS, protease inhibitors (Sigma P8340), and phosphatase inhibitors (Sigma P0044 and Sigma P5726). The cell assay plate with lysis buffer was shaken for 10 min at 4 °C and then 100 μl of cell lysate from each well was transferred to a goat anti-rabbit 96-well ELISA plate (Pierce #15135), which was previously incubated with rabbit anti-ABL1 antibody (Cell Signaling Technology #2862) diluted 1:200 in blocking buffer (Pierce StartingBlock). The cell lysate was incubated with the anti-c-ABL1-coated ELISA plate for 1 h at room temperature and then washed four times with Cell Signaling Technology ELISA Wash Buffer (from kit #7903). The final wash was removed by inverting the plate. One hundred microlitres of mouse monoclonal (IgG2b) anti-phospho-Tyr antibody (Santa Cruz Biotechnology SC508 HRP) diluted 1:5,000 was added to each well. The ELISA plate was then incubated for 45 min at room temperature with 100 μl per well. The plate was washed four times as described above, the final wash removed, and 100 μl of TMB substrate (Santa Cruz Biotechnology SC286967) was added to each well. Absorbance was measured at 655 nm during colour development or the reaction stopped by adding 50–100 μl per well of 0.16 M sulfuric acid stop solution and read at 450 nm.

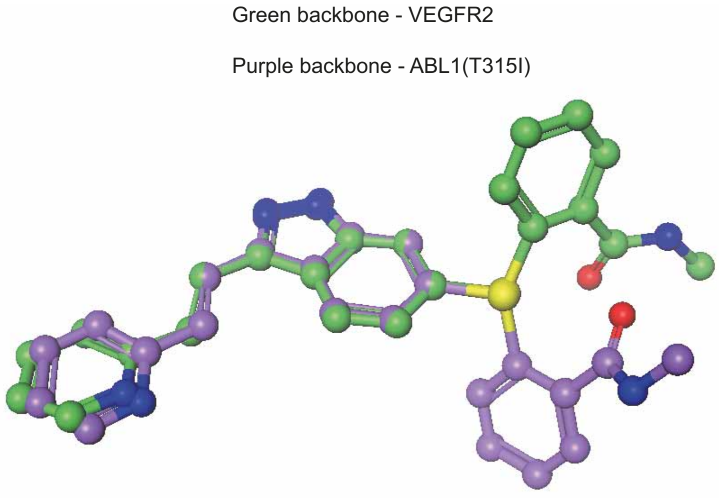
BCR–ABL Ba/F3 proliferation assay. For the proliferation assay, Ba/F3 cells were pipetted into a 50 ml tube, centrifuged at 180*g*, and the cell pellet was re-suspended in RPMI-1640 with 1% FBS, and 1% penicillin–streptomycin. The cells were counted using an Innovatis Cedex Cell Counter and seeded into a 96-well flat-bottom plate at 1,500 cells per well. Compounds were serially diluted in 100% DMSO as described above and then diluted 40-fold into RPMI-1640 with 1% FBS and 1% penicillin–streptomycin to yield a 5× source plate. Twenty-five microlitres was transferred from the 5× source plate to the cell assay plate and the cells incubated with test compounds for 4 days at 37 °C, 5% CO_2 95% air. On day 4 post drug addition, the cell assay plate was centrifuged at 180*g* for 2 min, 80 μl of supernatant was removed from each well, and 100 μl of fresh medium was added to each well. Fifteen microlitres of 1 mg ml^{-1} Resazurin (Sigma R7017) was then added to each well and the cell assay plate was incubated for 6 h at 37 °C, 5% CO_2 95% air. The fluorescence signal was read using 530 nm excitation and 595 nm emission wavelengths.

Engineered Ba/F3 cell proliferation method used in Extended Data Fig. 1. Cells were seeded at a concentration of 10,000 cells per well (10% FBS) in 96-well round-bottom cell culture plates with complete medium and in the presence of increasing concentrations of axitinib (range 0–10 μM). Cell proliferation was measured at 72 h using the tritiated thymidine incorporation assay as described previously[37]. The labelling time was for 8 h (started 64 h after seeding). Each test was performed in quadruplicate and replicated at least twice. Calculation of IC_{50} values was performed using GraphPad Prism software.

CRKL phosphorylation in CML patient cells. Mononuclear cells (5×10^6 per condition) of a CML patient sample (FHRB.1408) harbouring the T315I mutation (confirmed by reverse-transcription PCR and sequencing) were cultured overnight in either complete medium plus 0.1% DMSO or increasing concentrations of axitinib (1–1,000 nM; tenfold dilutions). Following incubation the cells were centrifuged, washed with cold PBS and lysed in 4% SDS, 0.1 M DTT and 0.1 M Tris. Lysate proteins were separated by SDS PAGE and transferred to an Immobilon-FL polyvinylidene difluoride (PVDF) membrane (Millipore). The membrane was blocked with 5% bovine serum albumin for 1 h, and incubated with phospho-CRKL (Y207; Cell Signaling Technologies (3181S); 1:1,000 dilution) and α-tubulin (Sigma Aldrich; T9026 mouse mAb; 1:1,000 dilution) antibodies in 5% BSA Tris-buffered saline and 0.05% Tween 20 (TBS-T) overnight at 4 °C. The membranes were then incubated with secondary infrared-labelled antibodies anti-mouse IRDye 680 and anti-rabbit IRDye 800CW (Odyssey; LI-COR Biosciences; 1:15,000 dilution) for 1 h at room temperature. The protein bands were visualized with the Odyssey imaging system (LI-COR Biosciences).

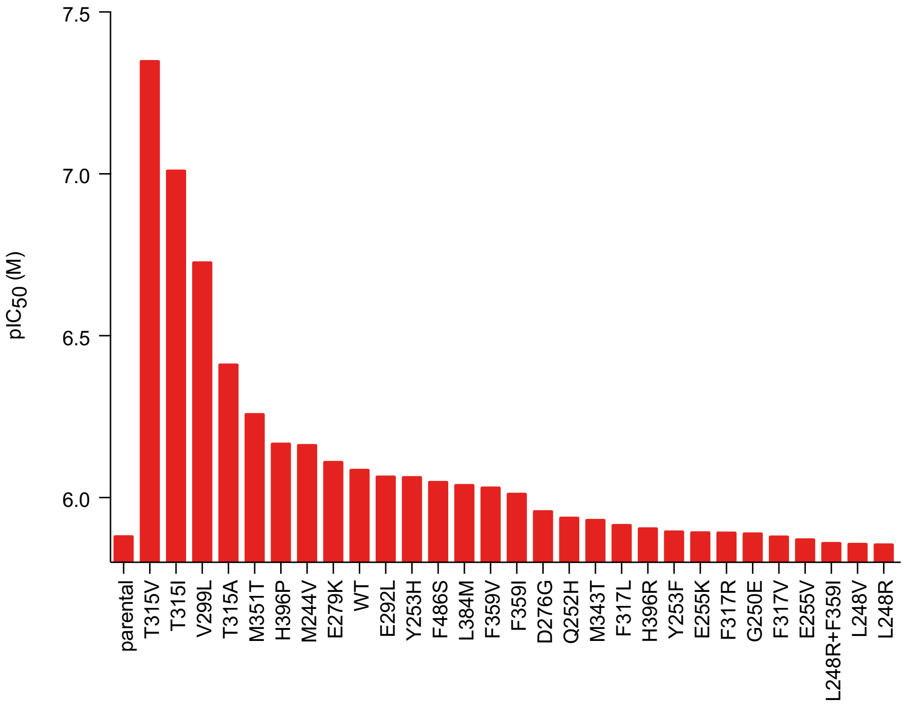
BCR–ABL1(T315I) transcript quantification. The proportion of T315I-mutation-positive transcripts was assessed with RT–qPCR using a mutation-specific forward primer, a reverse primer and a fluorescent TaqMan probe. The patients' bone marrow complementary DNA sample, that was found T315I-mutation-positive by Sanger sequencing, served as quantification standard for the follow-up samples. The cDNA from the T315I-positive sample was diluted into a negative control cDNA in a log-linear fashion to construct a dilution series and a corresponding standard curve. The standard curve was used for quantifying the T315I level in the samples taken at treatment day 0 and at day 14 (Fig. 4e). The results were normalized using *GUS* reference gene to compensate the differences in the RNA quality and the cDNA synthesis. Based on repeated measurements of the quantification standard in four independent runs, including two replicate analyses in each run, the between-run coefficient of variation for the measurement of the T315I:GUS ratio was 19.2%. The treatment time point samples were each measured by using three replicate analyses and an unpaired two-tailed *t*-test with Welch's correction indicated significant difference between the samples ($P < 0.05$).

31. Morrison, J. F. Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. *Biochim. Biophys. Acta* **185,** 269–286 (1969).
32. Kapust, R. B. *et al.* Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14,** 993–1000 (2001).
33. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
34. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).
35. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255 (1997).
36. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).
37. le Coutre, P. *et al.* In vivo eradication of human BCR/ABL-positive leukemia cells with an ABL kinase inhibitor. *J. Natl. Cancer Inst.* **91,** 163–168 (1999).
38. Solowiej, J., Chen, J. H., Zou, H. Y., Grant, S. K. & Murray, B. W. Substrate-specific conformational regulation of the receptor tyrosine kinase VEGFR2 catalytic domain. *ACS Chem. Biol.* **8,** 978–986 (2013).



Extended Data Figure 1 | Axitinib adopts significantly different binding conformation in the active site of ABL1(T315I) relative to VEGFR2. Overlay of the bound conformation of axitinib from VEGFR2 (green) (PDB ID: 4AGC) and ABL1(T315I) (purple) co-crystal structures.

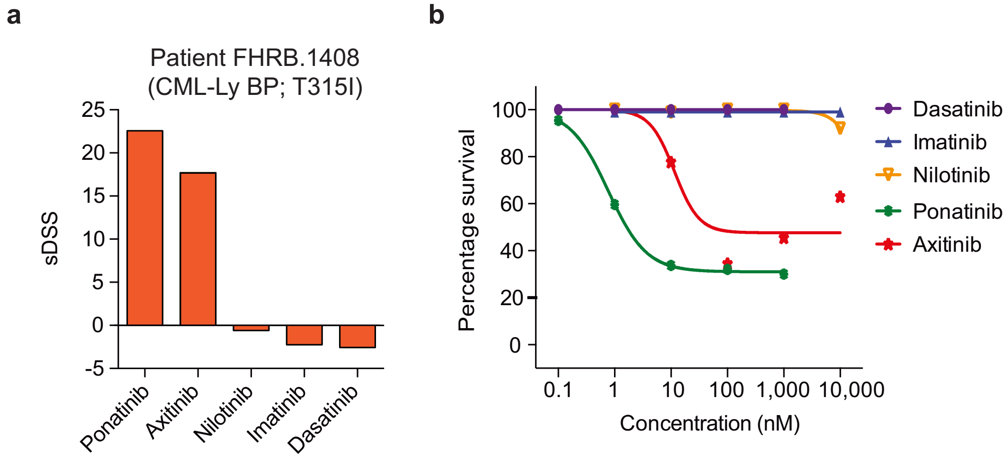
a



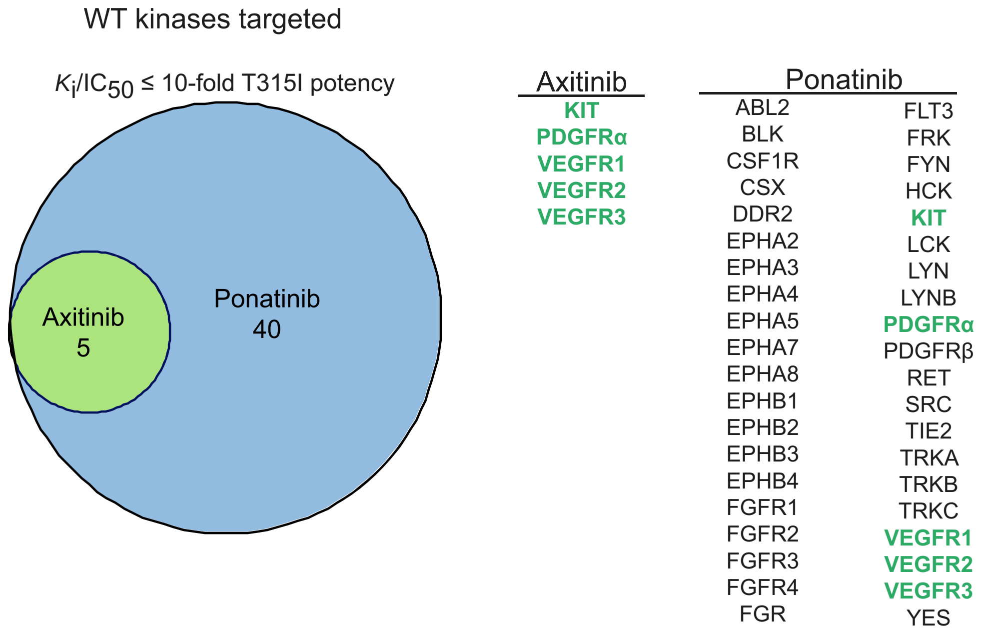
b

Ba/F3 cells	IC50 values (nM)
Parental	1319
T315V	45
T315I	98
V299L	188
T315A	389
M351T	553
H396P	683
M244V	690
E279K	778
WT	823
E292L	863
Y253H	867
F486S	897
L384M	916
F359V	933
F359I	975
D276G	1105
Q252H	1155
M343T	1175
F317L	1220
H396R	1247
Y253F	1275
E255K	1282
F317R	1286
G250E	1295
F317V	1320
E255V	1350
L248R+F359I	1385
L248V	1393
L248R	1399

Extended Data Figure 2 | Axitinib exhibits selective inhibition to BCR–ABL1 gatekeeper mutants in engineered Ba/F3 cells. a, Bar graph depicting *in vitro* cell proliferation data are displayed as pIC_{50} values. **b**, The corresponding IC_{50} values are shown in table format.



Extended Data Figure 3 | *Ex vivo* sensitivity to BCR–ABL1 inhibitors and axitinib in primary cells derived from a CML patient harbouring the T315I mutation (FHRB.1408). **a**, Waterfall plot showing the sDSS of axitinib, ponatinib, dasatinib, imatinib and nilotinib. **b**, Dose-response data of all approved BCR–ABL1 inhibitors and axitinib.



Extended Data Figure 4 | Comparison of publicly available target specificity profiles of axitinib[14,38] and ponatinib[20]. The target specificity profiles were evaluated at K_i/IC_{50} up to tenfold ABL1(T315I) potency (K_i = 0.1 nM axitinib; IC_{50} = 2 nM ponatinib) illustrating that axitinib is a much less promiscuous inhibitor than ponatinib and likely to have a better safety profile in Ph^+ leukaemia patients. Green labelled kinases are targeted by both inhibitors.

Extended Data Table 1 | X-ray data collection and refinement statistics

	ABL1(T315I) - axitinib	ABL1(WT) - axitinib
Data collection		
Space group	P 2 21 21	P 2 21 21
Cell dimensions		
a, *b*, *c* (Å)	56.818, 113.801, 131.164	56.626, 111.806, 128.881
a, *b*, *g* (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	65.6 – 2.4 (2.53-2.4) *	111.81-2.20 (2.32-2.20)
R_{sum} or R_{merge}	0.055 (0.624)	0.057 (0.489)
I/*sI*	12.2 (1.2)	11.5 (1.6)
Completeness (%)	99.9 (100.0)	98.5 (99.7)
Redundancy	6.5 (6.8)	4.8 (4.8)
Refinement		
Resolution (Å)	52.14-2.4 (2.47-2.40)	84.45-2.20 (2.25-2.20)
No. reflections	33967 (2680)	41498 (2630)
R_{work}/R_{free}	0.212/0.241	0.173/0.212
No. atoms	4617	4556
Protein	4365	4195
Ligand/ion	63	56
Water	189	305
B-factors		
Protein	58.37	42.52
Ligand/ion	55.69	38.77
Water	50.65	46.66
R.m.s deviations		
Bond lengths (Å)	0.009	0.010
Bond angles (°)	1.126	1.147

No. of crystals, 2.
* Highest resolution shell is shown in parenthesis.

Extended Data Table 2 | Clinical characteristics of patients

Patient	Age at diagnosis	Sex	Diagnosis	Karyotype at diagnosis	Therapy prior sampling	Time from diagnosis to sampling (months)	BCR-ABL1 mutations at sampling
FHRB.155	60	M	Acute lymphoblastic leukaemia	Hyperdiploid 47, t(9;22)(q34;q11) Trisomy 8	Imatinib and R-Cpm-Dox-Dxm-Vcr x 2, R-Asp-Dau-Dxm-Vcr, Dxm-Mp-Mtx-Vcr Mit-Vbl-Dxm-Cpm-Tio R-Mtx-Dxm-Vbl-Mp Dasatinib	55	T315I
FHRB.410	52	M	Chronic myeloid leukaemia in blast crisis	t(9;22)(q34;q11)	Imatinib	3	No detected mutations
FHRB.542	47	M	Acute lymphoblastic leukaemia	Hyperdiploid 48-50 t(9;22)(q34;q11), Trisomy 7 and 15	Dasatinib and Cpm-Dox-Dxm-Vcr x 2, Asp-Dau-Dxm-Vcr, Dxm-Mp-Mtx-Vcr HSCT Dasatinib	19	T315I
FHRB.651	34	F	Chronic myeloid leukaemia in 2nd blast crisis	t(9;22)(q34;q11)	Imatinib Dasatinib Dasatinib-interferon Ponatinib	20	E255K, V299L
FHRB.1278	44	M	Acute lymphoblastic leukaemia	t(9;22)(q34;q11), trisomy 4, 8, 14, 16, 17, +X, +Y	Dasatinib and Mit-Eto-Cyt HSCT Dasatinib	8	T315I
FHRB.1408	35	M	Chronic myeloid leukaemia in blast crisis	(9;22)(q34;q11)	Imatinib Dasatinib Cpm-Dox-Dxm-Vcr x 2	2	T315I

M, male; F, female; Asp, asparginase; Cpm, cyclophosphamide; Dau, daunorubicin; Dox, doxorubicin; Dxm, dexamethasone; HSCT, haematopoietic stem cell transplantation (allogeneic); Mit, mitoxantrone; Mp, mercaptopurine; Mtx, methotrexate; R, rituximab; Tio, thioguanine; Vcr, vincristine.

Extended Data Table 3 | Drug sensitivity scores for axitinib in a panel of BCR–ABL1(T315I)-positive, CML, AML and healthy donor control samples

Patient #	DSS	Sample Type
2071	0	Control
800	0	AML
997	0	AML
393_3	0	AML
393_1	0	AML
1064	0	AML
1145_1	0	AML
1145_6	0	AML
313	0.6	AML
1534	0.9	Control
1067	1.4	Control
1993	1.5	AML
693	1.7	Control
801	1.9	AML
252_3	2.6	AML
1346	2.8	Control
2095	2.9	AML
1145_3	3.1	AML
542	3.5	Ph+ ALL (T315I)
250	3.7	AML
1054	3.8	Control
1280	4.4	AML
600_3	4.7	AML
2200	5.4	Control
718	5.5	AML
2080	5.7	AML
410	6.0	CML
1867	6.0	AML
1690	6.5	AML
4701001	6.6	AML
784_2	7.7	AML
651	7.9	CML
784_1	8.5	AML
600_1	8.6	AML
1497	8.7	AML
1132	8.8	AML
252_5	8.8	AML
252_2	11.3	AML
560_1	11.8	AML
155	12.3	Ph+ ALL (T315I)
252_1	12.4	AML
600_2	12.8	AML
560_9	13.1	AML
1408_1	20	CML (T315I)
1278	29.6	Ph+ ALL (T315I)

doi:10.1038/nature13999

Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch

Alaji Bah[1,2], Robert M. Vernon[1,2], Zeba Siddiqui[1], Mickaël Krzeminski[1,2], Ranjith Muhandiram[2,3], Charlie Zhao[1], Nahum Sonenberg[4], Lewis E. Kay[1,2,3,5] & Julie D. Forman-Kay[1,2]

Intrinsically disordered proteins play important roles in cell signalling, transcription, translation and cell cycle regulation[1,2]. Although they lack stable tertiary structure, many intrinsically disordered proteins undergo disorder-to-order transitions upon binding to partners[3,4]. Similarly, several folded proteins use regulated order-to-disorder transitions to mediate biological function[5,6]. In principle, the function of intrinsically disordered proteins may be controlled by post-translational modifications that lead to structural changes such as folding, although this has not been observed. Here we show that multisite phosphorylation induces folding of the intrinsically disordered 4E-BP2, the major neural isoform of the family of three mammalian proteins that bind eIF4E and suppress cap-dependent translation initiation. In its non-phosphorylated state, 4E-BP2 interacts tightly with eIF4E using both a canonical YXXXXLΦ motif (starting at Y54) that undergoes a disorder-to-helix transition upon binding and a dynamic secondary binding site[7–11]. We demonstrate that phosphorylation at T37 and T46 induces folding of residues P18–R62 of 4E-BP2 into a four-stranded β-domain that sequesters the helical YXXXXLΦ motif into a partly buried β-strand, blocking its accessibility to eIF4E. The folded state of pT37pT46 4E-BP2 is weakly stable, decreasing affinity by 100-fold and leading to an order-to-disorder transition upon binding to eIF4E, whereas fully phosphorylated 4E-BP2 is more stable, decreasing affinity by a factor of approximately 4,000. These results highlight stabilization of a phosphorylation-induced fold as the essential mechanism for phospho-regulation of the 4E-BP:eIF4E interaction and exemplify a new mode of biological regulation mediated by intrinsically disordered proteins.

Proteins exist in a continuum of states, from ensembles of interconverting conformers for intrinsically disordered proteins (IDPs) to stable structures for folded proteins[2,4,12,13]. Although many undergo subtle conformational rearrangements in response to biological signals, some undergo larger changes: that is, disorder-to-order or order-to-disorder transitions[4] that can be mediated by interactions with ligands. Post-translational modifications (PTMs) such as phosphorylation can also potentially induce such transitions. While phosphorylation can convert the folded oligomerization domain of nucleophosmin into a disordered monomer[14,15], PTMs have been generally known only to stabilize or destabilize individual secondary structural elements in IDPs[16–18]. Thus, PTM-mediated folding of a protein domain, as shown here for the phosphorylation-induced folding of 4E-BP2, is a novel regulatory mechanism for IDPs.

Here we study the phospho-regulation of the 4E-BP:eIF4E interaction[19]. The 4E-BPs compete with eIF4G for eIF4E binding to prevent cap-dependent translation initiation by using similar canonical eIF4E-binding YXXXXLΦ motifs[20,21]. Non-phosphorylated or minimally phosphorylated 4E-BPs interact tightly with eIF4E, while the binding of highly phosphorylated 4E-BPs is much weaker and can be outcompeted by eIF4G. T37 and T46 are known to be phosphorylated first, followed by T70 and S65 (ref. 22). Interestingly, in their free states, highly phosphorylated 4E-BPs are very stable while non- or minimally phosphorylated 4E-BPs are targeted for degradation with ubiquitination at K57 (within the YXXXXLΦ motif) by the KLHL25-CUL3 E3 ligase[23]. However, there is no consensus on how phosphorylation regulates binding to eIF4E or affects the stability of 4E-BPs. Phosphorylation may result in electrostatic repulsion with the negative surface of eIF4E[21] or S65 phosphorylation could inhibit binding by destabilizing the YXXXXLΦ motif helical structure[24]. The inability of glutamic acid phospho-mimetics to weaken 4E-BP:eIF4E binding has also suggested that additional PTMs may be required[25].

To address the phospho-regulatory mechanism, we have used NMR, isothermal titration calorimetry (ITC) and mutagenesis. Figure 1 and Extended Data Fig. 1 show structural and dynamic properties of wild-type (WT) 4E-BP2 uniformly phosphorylated at T37, T46, S65, T70 and S83 (Fig. 1a). IDPs such as non-phosphorylated 4E-BP2 have intense peaks with narrow ^{1}HN chemical shift dispersion[26]. Unlike other IDPs for which phosphorylation only causes downfield chemical shifts of phosphorylated residues[27], 4E-BP2 phosphorylation induces widespread downfield and upfield chemical shifts for residues spanning T19–R62, suggesting folding upon phosphorylation (Fig. 1b). Notably, G39 and G48, the first glycines in the six-residue repeat sequences (TTPGGT) containing two of the five phosphorylation sites, show dramatic downfield ^{1}HN chemical shift changes (Fig. 1b, inset). Peaks for the rest of the protein are intense with narrow ^{1}HN chemical shift dispersion, indicating that these residues remain disordered. Like its non-phosphorylated state[11], phosphorylated 4E-BP2 exchanges between major and minor conformers, probably from *cis–trans* isomerization of the multiple prolines and between unfolded and partly folded states for the folded region (see below). Lower [^{1}H]–^{15}N nuclear Overhauser effect (NOE) values are expected for IDPs because of their rapid motions[26,28]. Elevated [^{1}H]–^{15}N NOEs for most residues from T19–R62 (Fig. 1c), along with the large increase in chemical shift dispersion and the high amide proton temperature coefficients indicative of intramolecular hydrogen bonding[29] (Extended Data Fig. 1d), provide strong evidence that this region folds upon phosphorylation, while the rest of the protein remains disordered.

We made alanine mutations to mimic *in vivo* phosphorylation states of 4E-BP2 (ref. 22) (Extended Data Fig. 2). No significant change in global dispersion was observed for 4E-BP2 phosphorylated only at S65/T70/S83 (Extended Data Fig. 2a), demonstrating that it remains disordered, while phosphorylating T37 and T46 (pT37pT46) induces a 4E-BP2 fold identical to phosphorylated wild type (pWT) (Extended Data Fig. 2b). Interestingly, when phosphorylated individually, pT37 or pT46 result in a partly folded state, with some chemical shift changes indicative of ordered structure (pT37) and the presence of one β-turn leading to one characteristic downfield shifted glycine peak (G39 for pT37, G48 for pT46; Extended Data Fig. 2c, d). Thus, phosphorylation of both T37 and T46 is necessary and sufficient for phosphorylation-induced folding of 4E-BP2.

We determined the structure of P18–R62 using CS-Rosetta[30] including initially only chemical shifts, and subsequently adding NOEs (see

[1]Molecular Structure and Function Program, Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada. [2]Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [3]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [4]Department of Biochemistry and Goodman Cancer Research Centre, McGill University, Montréal, Quebec H3G 1Y6, Canada. [5]Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada.

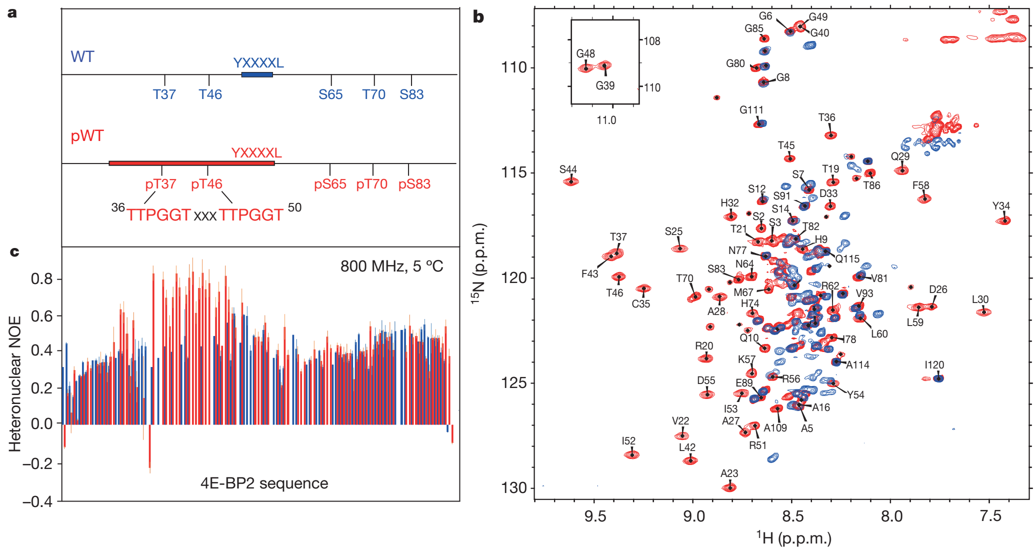


Figure 1 | Effects of phosphorylation on the structural and dynamic properties of 4E-BP2. **a**, Schematic representation of 4E-BP2 showing the relative positions of the phosphorylation sites, the canonical eIF4E binding site (thick blue bar) and the region which undergoes phosphorylation-induced folding (thick red bar). **b**, **c**, Overlay of (**b**) ^{1}H–^{15}N heteronuclear single quantum correlation (HSQC) spectra and (**c**) ^{1}H–^{15}N NOE values at 800 MHz, 5 °C for non-phosphorylated WT (blue) and phosphorylated WT (red), respectively. Missing data represent prolines and residues that are too overlapped/weak to be accurately quantified. Errors (standard deviations) around the average values are based on multiple repeats ($n = 3$) of the experiment.

Methods, Extended Data Figs 3–6 and Extended Data Table 1a, b). The domain is a four-stranded β-fold (Fig. 2), with β1(T19–I24), β2(C35–pT37), β3(L42–T45), β4(R51–R56) and a 3_{10} helix (A27–Q29). pT37 and pT46 are central to a network of hydrogen bonds that stabilize tight β-turns composed of the identical pTPGGT motifs connecting strands β2 with β3 (pT37–T41) and β3 with β4 (pT46–T50). The hydrogen bonding correlates well with ^{1}HN temperature coefficients (Extended Data Fig. 1d). G39 and G48 amide protons hydrogen bond to T37 and T46 phosphate groups, respectively, explaining their unexpectedly large downfield chemical shifts (Fig. 1a, inset, and Extended Data Fig. 2b). The YXXXXLϕ eIF4E-binding motif forms part of strand β4 with Y54 largely buried within a hydrophobic cluster also involving V22, I24, L30, P31, Y34 and I52 (Fig. 2a). The phosphorylation-induced structure provides a mechanism by which phosphorylation reduces eIF4E binding by sequestering the helical YXXXXLϕ motif into a β-strand. Its burial stabilizes the β-fold; phosphorylated Y54A/L59A [p(Y54A/L59A)] retains the β-turns, evident from the characteristic downfield ^{1}HN shifts of G39 and G48, but no longer folds (Fig. 3a). A stable fold sequestering YXXXXLϕ could also block ubiquitination of K57, thereby preventing degradation *in vivo*[23].

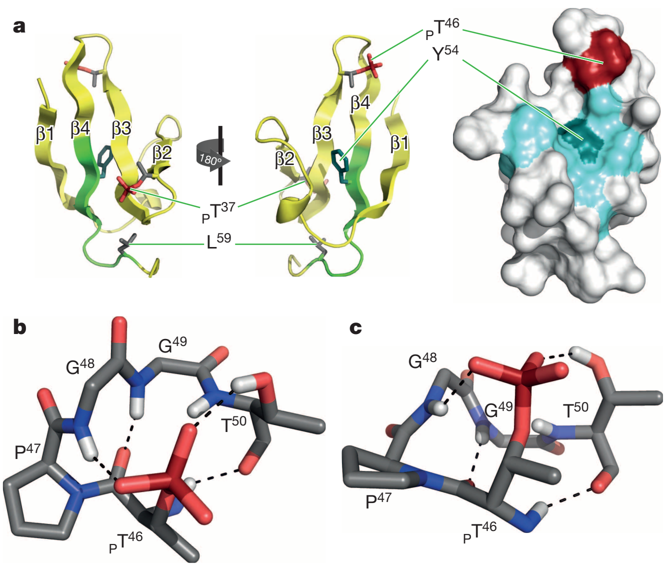


Figure 2 | Phosphorylation-induced structure of the major state of residues R18–R62 of 4E-BP2. **a**, Cartoon (left) and surface (right) representations of the solution NMR structure. Phosphorylated residues, pT37 and pT46 (red), the surface formed by residues of the hydrophobic cluster (right, cyan) and Y54 (cyan stick representation (left) or dark cyan surface (right)) are shown. The residues in the canonical eIF4E binding site (YXXXXLΦ) are shown in green (left), demonstrating the binding-incompatible extended structure. **b**, **c**, Hydrogen-bonding networks formed by the pTPGGT motifs. Note that pT46–T50 is shown in two different views and the same hydrogen bonds are formed for pT37–T41.

ITC (Extended Data Table 2 and Extended Data Fig. 7) and NMR data were obtained on 4E-BP2 variants to probe the role of electrostatics in folding and reducing eIF4E affinity. Unlike pT37pT46, neither T37D/T46D nor T37E/T46E (phospho-mimics) showed evidence of folding (Extended Data Fig. 2e, f). In contrast to pT37pT46, which has reduced affinity (dissociation constant (K_d) = 267 ± 32 nM), T37D/T46D and T37E/T46E bind tightly to eIF4E (K_d = 3.89 ± 1.1 nM and 4.37 ± 0.8 nM, respectively), similar to non-phosphorylated 4E-BP2 (K_d = 3.20 ± 0.6 nM). Thus acidic residues do not mimic phosphorylation by inducing folding of 4E-BP2 or reducing eIF4E affinity. The binding affinity of pS65pT70pS83 4E-BP2 for eIF4E was very tight (K_d = 11.3 ± 2.9 nM), as expected since a folded structure is not formed (Extended Data Fig. 2a). Strikingly, five-phospho protein (pWT) decreases affinity by about three orders of magnitude (K_d = 12,320 ± 600 nM). Valines were substituted for the first glycines in the TPGGT motifs (G39V/G48V) of 4E-BP2 to disrupt hairpin formation through steric contacts and prevent folding. As predicted from ΔΔG calculations (Extended Data Table 1c), folding was not induced when G39V/G48V was fully phosphorylated at all five sites [p(G39V/G48V)] (Fig. 3b). Notably, the affinity of the p(G39V/G48V) for eIF4E is high (K_d = 36.1 ± 3.5 nM), only an order of magnitude weaker than that of non-phosphorylated 4E-BP2 (K_d = 3.20 ± 0.6 nM) and approximately 2.5 orders of magnitude stronger than pWT (K_d = 12,320 ± 600 nM). These data, together with unfolding energies of pT37pT46 (~2.6 kcal mol^{-1}) and pWT (~4.8 kcal mol^{-1}) estimated from K_d values (Extended Data Table 2), suggest that, although phosphorylation of only T37 and T46 is required to induce folding, phosphorylation of S65, T70 and S83 stabilizes the fold. This is possibly through long-range transient electrostatic

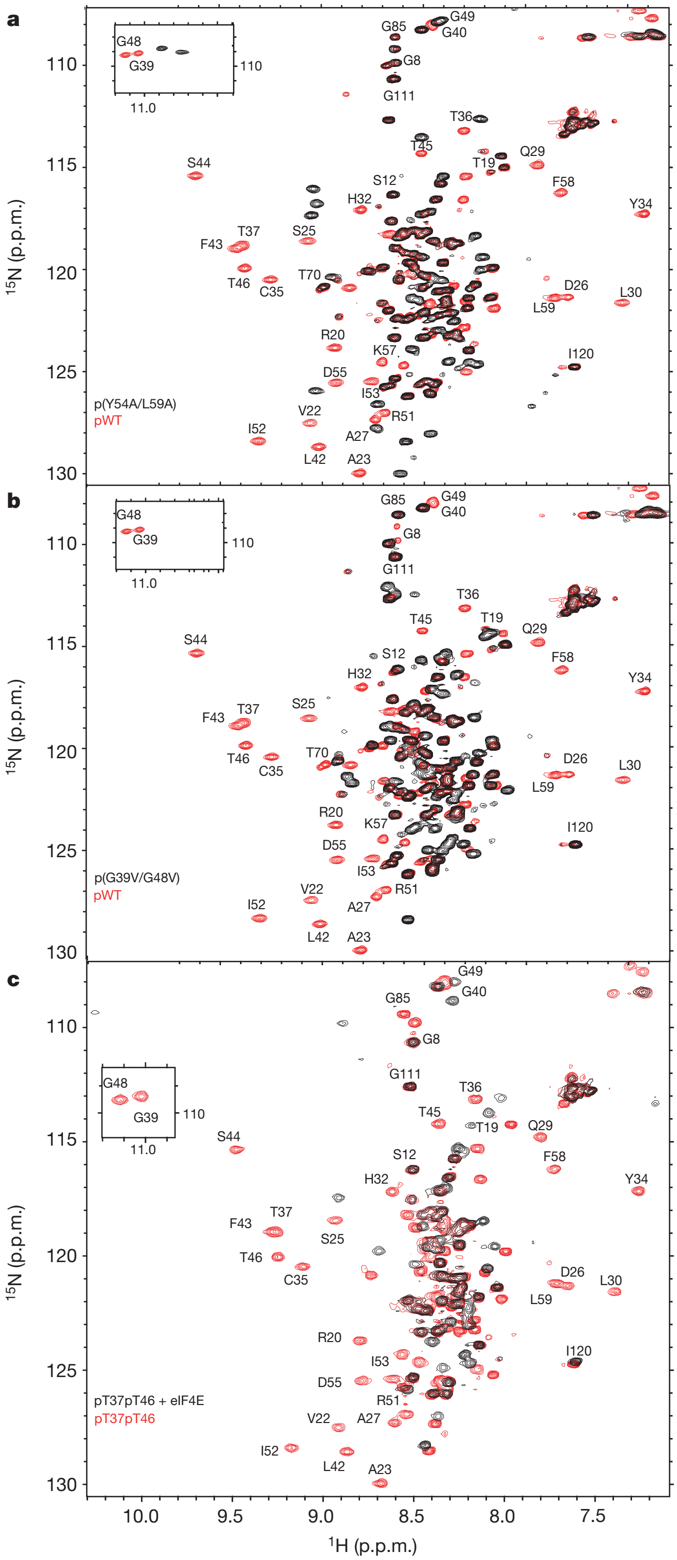


Figure 3 | Probing the structural and binding properties of phosphorylated 4E-BP2. **a**, **b**, ^{1}H–^{15}N HSQC spectra of pWT (red) overlaid with (**a**) p(Y54A/L59A) and (**b**) p(G39V/G48V). **c**, Spectrum of pT37pT46 in isolated (red) and eIF4E-bound (black) states, demonstrating an order-to-disorder transition upon eIF4E binding.

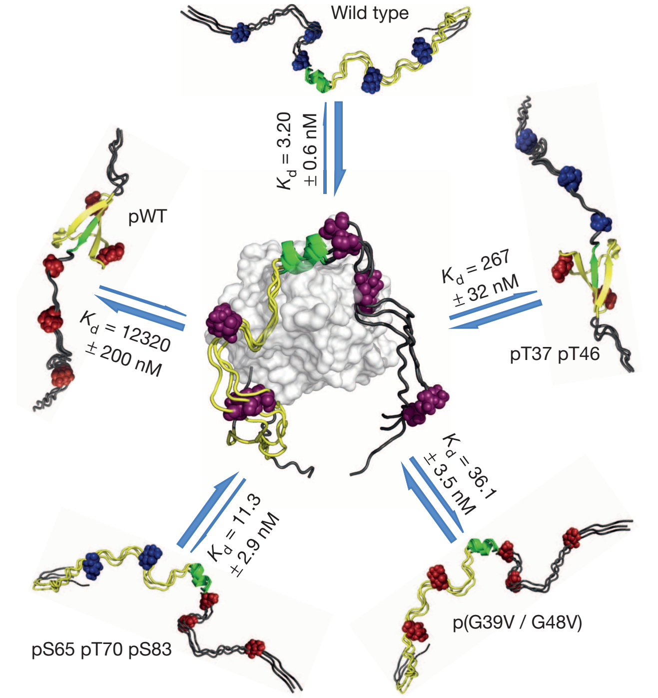


Figure 4 | Phospho-regulation of the eIF4E:4E-BP interaction. Schematic representations of different phosphorylation states of 4E-BP2 and K_d values for eIF4E interaction (grey surface). The 4E-BP2 residues spanning Q10-D90 are shown with the canonical eIF4E binding site (YXXXXLΦ, green), other residues involved in the phosphorylation-induced fold (P18–R62, yellow) and non-phosphorylated (blue) and phosphorylated (red) Ser/Thr residues. Upon phosphorylation, 4E-BP2 undergoes a disorder-to-order transition, which significantly weakens eIF4E interaction, allowing eIF4E to bind eIF4G to initiate translation. Three representations, without spanning the likely broad conformational space sampled, are shown to indicate the disorder present in 4E-BP2 regions that are not folded.

interactions between the phosphorylated acidic carboxy (C) terminus with the basic folded domain. Sequestration of the canonical eIF4E motif by pT37pT46 reduces affinity by about two orders of magnitude, and not the three orders observed for the fully phosphorylated form, possibly because of a binding-competent minor disordered state that is visible in the spectrum, providing additional evidence for lower stability of the pT37pT46 folded state. These affinities and calculations support the view that stabilization of a phosphorylation-induced folded structure plays a dominant role in weakening the eIF4E:4E-BP2 interaction, with a probable small contribution from electrostatic repulsion from eIF4E as reflected in the ten-fold reduction in affinity for p(G39V/G48V) from non-phosphorylated WT 4E-BP2.

Because the 4E-BP2 β-fold sequesters the eIF4E-binding interface, binding to eIF4E must be coupled to unfolding. Although it is folded in the absence of target, pT37pT46 undergoes an order-to-disorder transition upon binding to eIF4E (Fig. 3c), as established by the resulting poorly resolved spectrum confined to a narrow amide proton chemical shift range and by the disappearance of well dispersed folded peaks of the *apo*-state, except for a single very weak ^{1}HN glycine peak at 10.25 p.p.m. that may reflect a low population of a β-turn (Fig. 3c and Extended Data Fig. 2g). The spectrum is similar to that of non-phosphorylated 4E-BP2 (ref. 11); both complexes are disordered with significant chemical shift differences (Extended Data Fig. 2g), probably reflecting changes in interactions due to different binding affinities (3.20 ± 0.6 nM versus 267 ± 32 nM) and effects of phosphates, including low population of a β-turn.

Our study provides key insights into how the structural polymorphism of 4E-BP2 allows it to regulate translation initiation through PTM-mediated folding (Fig. 4). This mechanism establishes a new means of IDP-mediated control of biological function. Large structural changes such as folding could sequester or enhance the accessibilities of protein binding and other PTM sites or provide new interaction surfaces, thereby expanding signalling output. Importantly, PTM-induced folding has important potential impact for targeting IDP interactions for therapeutics.

Since the stability of the phosphorylation-induced 4E-BP fold is critical in controlling binding to eIF4E, small molecules that stabilize or destabilize folding of 4E-BPs are likely to be potent modulators of translation initiation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 April; accepted 24 October 2014.
Published online 22 December 2014.

1. Dunker, A. K. & Uversky, V. N. Signal transduction via unstructured protein conduits. *Nature Chem. Biol.* **4,** 229–230 (2008).
2. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature Rev. Mol. Cell Biol.* **6,** 197–208 (2005).
3. Wright, P. E. & Dyson, H. J. Linking folding and binding. *Curr. Opin. Struct. Biol.* **19,** 31–38 (2009).
4. Uversky, V. N. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta* **1834,** 932–951 (2013).
5. Mitrea, D. M. & Kriwacki, R. W. Regulated unfolding of proteins in signaling. *FEBS Lett.* **587,** 1081–1088 (2013).
6. Schultz, J. E. & Natarajan, J. Regulated unfolding: a basic principle of intraprotein signaling in modular proteins. *Trends Biochem. Sci.* **38,** 538–545 (2013).
7. Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136,** 731–745 (2009).
8. Tomoo, K., Abiko, F., Miyagawa, H., Kitamura, K. & Ishida, T. Effect of N-terminal region of eIF4E and Ser65-phosphorylation of 4E-BP1 on interaction between eIF4E and 4E-BP1 fragment peptide. *J. Biochem.* **140,** 237–246 (2006).
9. Fletcher, C. M. *et al.* 4E binding proteins inhibit the translation factor eIF4E without folded structure. *Biochemistry* **37,** 9–15 (1998).
10. Gosselin, P. *et al.* The translational repressor 4E-BP called to order by eIF4E: new structural insights by SAXS. *Nucleic Acids Res.* **39,** 3496–3503 (2011).
11. Lukhele, S., Bah, A., Lin, H., Sonenberg, N. & Forman-Kay, J. D. Interaction of the eukaryotic initiation factor 4E with 4E-BP2 at a dynamic bipartite interface. *Structure* **21,** 1–11 (2013).
12. Mittag, T., Kay, L. E. & Forman-Kay, J. D. Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* **23,** 105–116 (2010).
13. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313,** 1638–1642 (2006).
14. Mitrea, D. M. & Kriwacki, R. W. Cryptic disorder: an order-disorder transformation regulates the function of nucleophosmin. *Pac. Symp. Biocomput.* **2012,** 152–163 (2012).
15. Mitrea, D. M. *et al.* Structural polymorphism in the N-terminal oligomerization domain of NPM1. *Proc. Natl Acad. Sci. USA* **111,** 4466–4471 (2014).
16. Pufall, M. A. *et al.* Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* **309,** 142–145 (2005).
17. Theillet, F. X. *et al.* Cell signaling, post-translational protein modifications and NMR spectroscopy. *J. Biomol. NMR* **54,** 217–236 (2012).
18. Espinoza-Fonseca, L. M., Kast, D. & Thomas, D. D. Thermodynamic and structural basis of phosphorylation-induced disorder-to-order transition in the regulatory light chain of smooth muscle myosin. *J. Am. Chem. Soc.* **130,** 12208–12209 (2008).
19. Lin, T.-A. *et al.* PHAS-I as a link between mitogen-activated protein kinase and translation initiation. *Science* **266,** 653–656 (1994).
20. Mader, S., Lee, H., Pause, A. & Sonenberg, N. The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. *Mol. Cell. Biol.* **15,** 4990–4997 (1995).
21. Marcotrigiano, J., Gingras, A.-C., Sonenberg, N. & Burley, S. K. Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. *Mol. Cell* **3,** 707–716 (1999).
22. Gingras, A.-C. *et al.* Regulation of 4E-BP1 phosphorylation: a novel two step mechanism. *Genes Dev.* **13,** 1422–1437 (1999).
23. Yanagiya, A. *et al.* Translational homeostasis via the mRNA cap-binding protein, eIF4E. *Mol. Cell* **46,** 847–858 (2012).
24. Tait, S. *et al.* Local control of a disorder-order transition in 4E-BP1 underpins regulation of translation via eIF4E. *Proc. Natl Acad. Sci. USA* **107,** 17627–17632 (2010).
25. Oulhen, N. *et al.* A variant mimicking hyperphosphorylated 4E-BP inhibits protein synthesis in a sea urchin cell-free, cap-dependent translation system. *PLoS ONE* **4,** e5070 (2009).
26. Dyson, H. J. & Wright, P. E. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104,** 3607–3622 (2004).
27. Selenko, P. *et al.* *In situ* observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nature Struct. Mol. Biol.* **15,** 321–329 (2008).
28. Klein-Seetharaman, J. *et al.* Long-range interactions within a nonnative protein. *Science* **295,** 1719–1722 (2002).
29. Cierpicki, T. & Otlewski, J. Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J. Biomol. NMR* **21,** 249–261 (2001).
30. Vernon, R., Shen, Y., Baker, D. & Lange, O. F. Improved chemical shift based fragment selection for CS-Rosetta using Rosetta3 fragment picker. *J. Biomol. NMR* **57,** 117–127 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Augustyniak, Z. Bozoky, V. Csizmok, J. Dawson and P. Farber for discussions, and A. Hansen for help in NMR data processing. A. Chong, R. Hudson and H. Lin are acknowledged for their technical expertise. This work was funded by grants from the Canadian Institutes of Health Research (MOP-114985, MOP-119579) and the Canadian Cancer Society to J.D.F.-K. A.B. was partly supported by a Restracomp award from the Hospital for Sick Children and a post-doctoral fellowship from the Canadian Institutes of Health Research (CIHR). R.M.V. was partly supported by a post-doctoral fellowship from the CIHR Strategic Training Program in Protein Folding and Interaction Dynamics. Z.S. was partly supported by the Summer Research Program at the Hospital for Sick Children.

Author Contributions A.B. and J.D.F.-K. designed experiments. A.B., Z.S. and N.S. contributed reagents. A.B., L.E.K. and R.M. performed NMR experiments. R.M.V. and M.K. performed structure calculations. A.B., Z.S. and C.Z. performed biochemical experiments. A.B., R.M.V., L.E.K. and J.D.F.-K. analysed data. A.B., R.M.V., M.K., N.S., L.E.K. and J.D.F.-K. wrote and edited the paper.

Author Information Chemical shifts of non- and fully phosphorylated 4E-BP2 have been deposited in the Biological Magnetic Resonance Bank (BMRB) under accession numbers 19114 (ref. 11) and 19905, respectively. The coordinates for the folded state of phosphorylated 4E-BP2 has been deposited in the Protein Data Bank (accession number 2MX4). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.F.-K. (forman@sickkids.ca).

METHODS

Protein expression and purification. Small ubiquitin-like modifier (Sumo) fusion constructs of human eIF4E or wild type/mutant 4E-BP2 were expressed and purified to homogeneity as previously described[11]. Expression and purification of activated His-tagged Erk2 used a protocol and plasmid co-expressing Erk2 and MEK1 obtained from Attila Remenyi at Eötvös Loránd University.

Phosphorylation of 4E-BP2. All 4E-BP2 constructs were phosphorylated to homogeneity with Erk2 using a dialysis technique. Briefly, each phosphorylation reaction was made up of ~50 ml of phosphorylation buffer (50 mM Tris pH 7.5 at 25 °C, 1 mM EGTA, 2 mM DTT, 20 mM $MgCl_2$ and 10 mM ATP) containing ~20 μM 4E-BP2 and ~5 μM Erk2 in a dialysis bag placed in 1 l of phosphorylation buffer. Phosphorylation was allowed to proceed overnight with stirring before a 20 μl aliquot of the reaction was removed for mass spectrometric analysis. Once the expected number of sites was uniformly phosphorylated, the reaction was stopped by running over a nickel-nitrilotriacetic acid (Ni-NTA) column to remove the kinase. Flow-through and wash fractions were collected, concentrated and then purified via reverse-phase high-performance liquid chromatography (HPLC). HPLC fractions containing phosphorylated protein were pooled, concentrated and dialysed in 4 l of buffer for about 16 h. Site-directed mutagenesis and the above-described method of phosphorylation by dialysis allowed the generation of reproducible large quantities of samples phosphorylated at any chosen combination of phosphorylation sites for biophysical studies (Extended Data Table 1). Mass spectrometry and NMR analysis of all samples confirmed the phosphorylation state (Extended Data Fig. 8).

NMR spectroscopy and binding studies. NMR samples comprised approximately 0.1–1.0 mM $^{1}H^{15}N^{13}C$-labelled protein in a buffer containing 30 mM Na_2HPO_4, 100 mM NaCl, 2 mM DTT, 1 mM EDTA, 10% D_2O v/v, pH 6.0. All NMR experiments were performed on Varian INOVA 500, 600 and 800 MHz spectrometers equipped with pulsed-field gradient units and triple resonance probes with a 600 MHz spectrometer equipped with a cryogenically cooled probe. NMR data sets were processed with the NMRPipe software package[31] and analysed using SPARKY[32] and FuDA (http://pound.med.utoronto.ca/~flemming/fuda/).

ITC studies and most NMR experiments, including those for chemical shift assignment, were performed at pH 6.0 and 20 °C. Temperature- and pH-dependent NMR experiments were recorded with temperature and pH values that varied from 5 to 35 °C and 4.0 to 10.3, respectively. The $[^{1}H]$–^{15}N-NOE relaxation measurements were performed at pH 6.0, 5 °C and 800 MHz. Other experimental details for the ITC and NMR were as previously described for the non-phosphorylated 4E-BP2 (ref. 11). To obtain distance restraints for CS-ROSETTA structural calculations (see below), we also recorded combined ^{15}N- and ^{13}C-edited nuclear Overhauser effect spectroscopy (CN-NOESY) and ^{15}N-edited NOESY[33] data sets, with amide proton temperature coefficients measured to identify amides involved in intramolecular hydrogen bonding[29].

Calculation of structures of the folded state of phosphorylated 4E-BP2 (pWT). The structure of the major state of pWT was calculated with the CS-Rosetta program[34] using chemical shifts with ^{1}H–^{1}H NOEs as distance restraints. First, TALOS+[35] was used to determine the secondary structure propensity of pWT on the basis of the measured backbone ^{1}H, ^{15}N and $^{13}C\alpha$, and $^{13}C\beta$ chemical shifts; a prediction of four strands and a small helix was obtained for residues P18–R62 (Extended Data Fig. 5). Interestingly, this region showed both well-dispersed ^{1}HN chemical shifts and intramolecular amide proton hydrogen-bond formation as well as high positive ^{1}H–^{15}N heteronuclear NOEs for most residues (see text, Fig. 1 and Extended Data Fig. 1d). To test whether this secondary structure was compatible with an independently folded domain, CS-Rosetta2 was used to generate approximately 20,000 models using the standard protocol and the best five structures by the Rosetta energy function converged to a single topology within 2.1 Å in Cα root mean square deviation (r.m.s.d.) for residues P18–K57, confirming that the chemical shift data were compatible with the formation of a folded domain (Extended Data Fig. 3). Note that ^{1}H–^{1}H NOEs involving residues C-terminal to K57 could not be confidently assigned, so structure calculations were performed only for residues P18–K57 and, thus, no structural data are available for the five motionally restricted residues F58–R62.

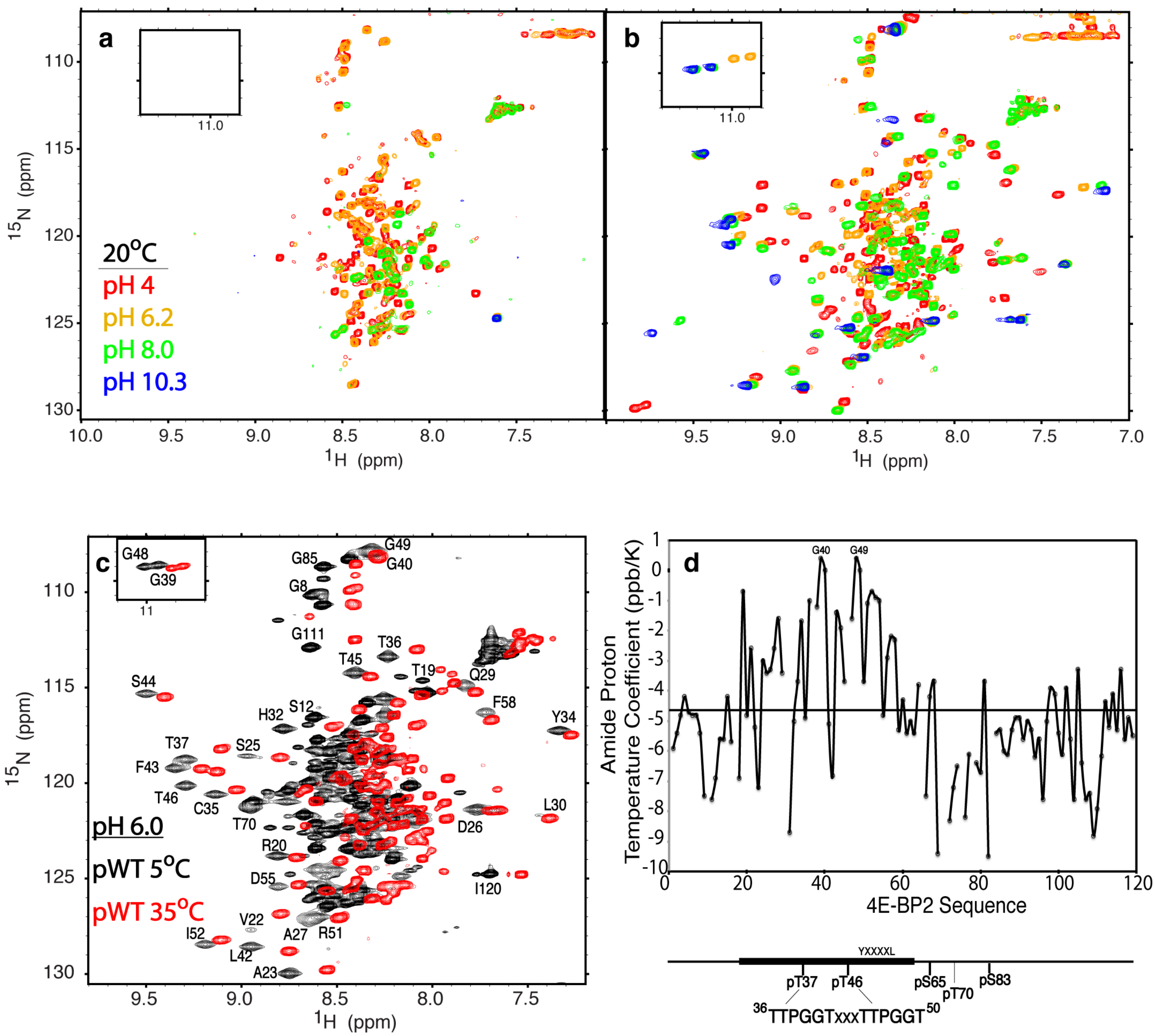
To provide further structural details for the folded state, ^{1}H–^{1}H NOE intensities were then used to identify distance restraints between atoms. Complications arose during the assignment of the NOESY data, as spectra showed significant overlap due to large regions of disorder in much of the protein as well as line broadening and evidence for minor states for the residues composing the fold (see text). Consequently preliminary attempts to assign the NOE data automatically failed, and manual assignment was complicated by ambiguity. We thus selected our most confident assignments for a final set of 494 discrete atom–atom pairs and six ambiguous pairs, which we then split into strong, medium and weak distance restraints by taking the highest 25%, middle 50% and lowest 25% by peak intensity, with separate boundaries defined for NOEs involving HN atoms and those that did not. Importantly, the CS-Rosetta2 ensemble using both chemical shifts and NOEs demonstrated the same overall fold as that calculated by CS-Rosetta2 using only chemical shifts. Comparing our distance restraints with the CS-Rosetta2 ensemble calculated with NOEs showed that the majority of the restraints were satisfied, with an average of 79.7% of the full set and 61.0% of the long-range contacts (atom pairs separated by ten or more residues) satisfied across the ensemble (Extended Data Table 1). Notably, the only clusters of long-range distance restraints observed in the NOEs corresponded to the strand pairs in the ensemble.

On the basis of our expectation that a significant number of the violated restraints come from transiently populated minor states, and to focus our calculations on the structure of the major state, we produced models using a protocol that allowed individual restraints to be violated while selecting for models that had the highest overall fit to the data. For this, we first converted the NOESY data into ambiguous distance restraints for use in CS-Rosetta2, then identified the largest violations by generating 2,000 models and removing 13 constraints from atom pairs that were not observed within 6/7/8 Å (for strong/medium/weak NOEs) of each other in the structure with the best constraint score. We then ran a second round of CS-Rosetta2 for a fixed amount of time, generating 20,359 models using the remaining restraints filtered down to 6,490 models by taking structures where the amide protons of G39 and G48 were both hydrogen bonded (as indicated by the temperature coefficients and large downfield ^{1}HN chemical shifts), then scored each model on the basis of their percentage of long-range NOEs satisfied (closest atom pair distances within 4/5/6 Å for strong/medium/weak). We then used those percentages to filter the pool further to 325 structures representing the best 5% by NOEs satisfied, and finally out of those selected the best 20 by their Rosetta energies (Extended Data Table 1 and Extended Data Figs 3 and 4).

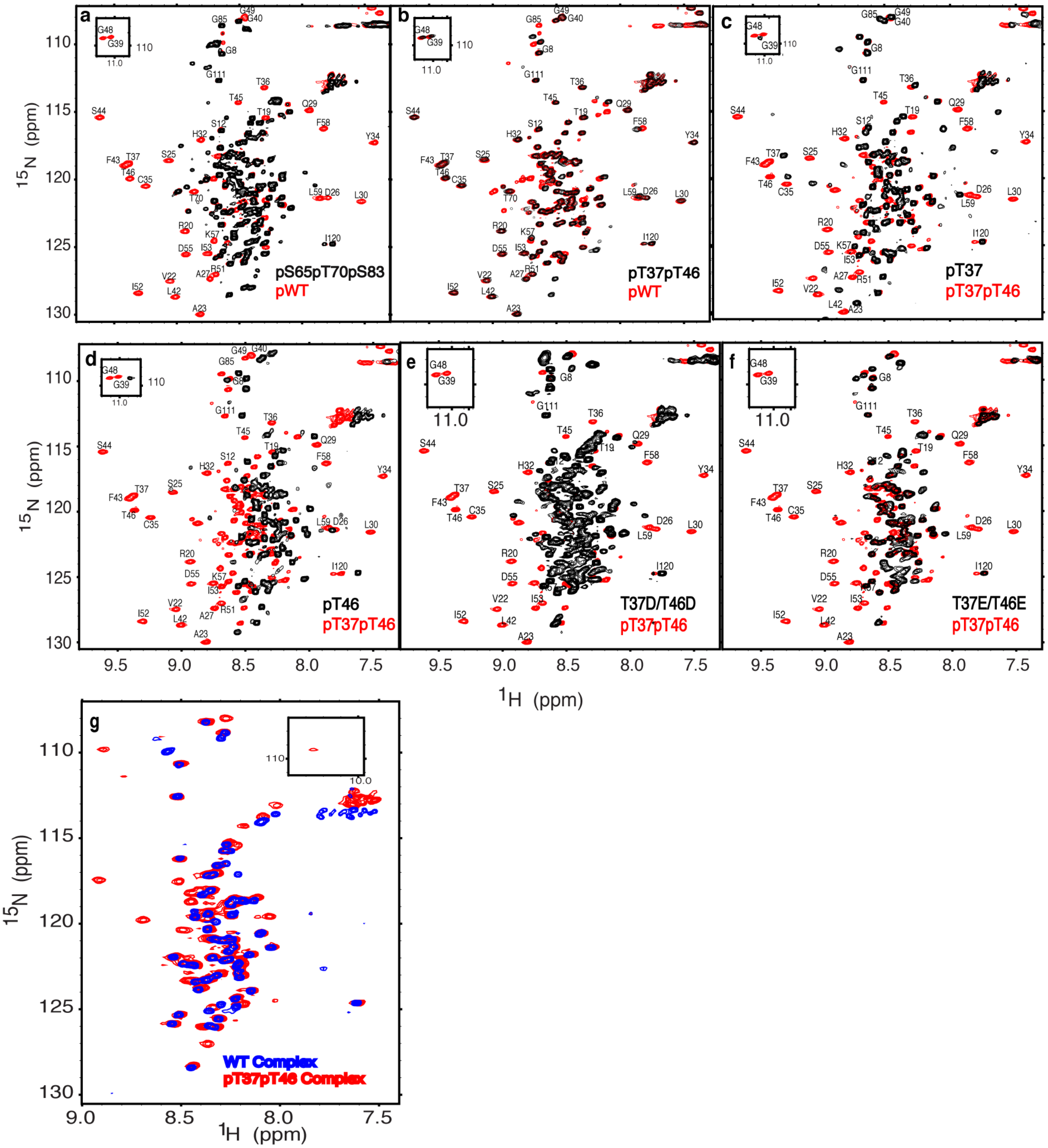
A contact map plotting NOE and NOE violations by their residue pairs (Extended Data Fig. 6a) shows how the folded topology emerges from the majority of the data, with violations clustering in specific locations. Consistent with our interpretation, violations appear to arise from conformational exchange with minor states that retain significant folded structure, representing folding intermediates or partial structures that have lost one or two contacts, as demonstrated by analysis of specific NOEs (Extended Data Fig. 6b–d). Note that these could give rise to chemical shifts that overlap the major folded state for most residues with additional resonances for the disordered portions, consistent with the observation of minor peaks in the disordered region of the spectrum.

Validating the structural model of the phosphorylation-induced folded domain. To test the four β-stranded structural model, we performed mutagenesis of strategic residues in the fold using both *in silico* ΔΔG predictions[36] and experimental site-directed mutagenesis (see text and Extended Data Table 1c). According to the ΔΔG predictions made on the basis of the lowest scoring CS-Rosetta structure before the incorporation of NOE-based restraints, perturbing the TpTPGGT motifs, which form the tight turns of the hairpin, or the canonical binding motif (YXXXXLΦ) significantly destabilizes the fold. Some of these mutations were tested experimentally and shown to destabilize the fold (Fig. 3, Extended Data Figs 2 and 7 and Extended Data Table 2).

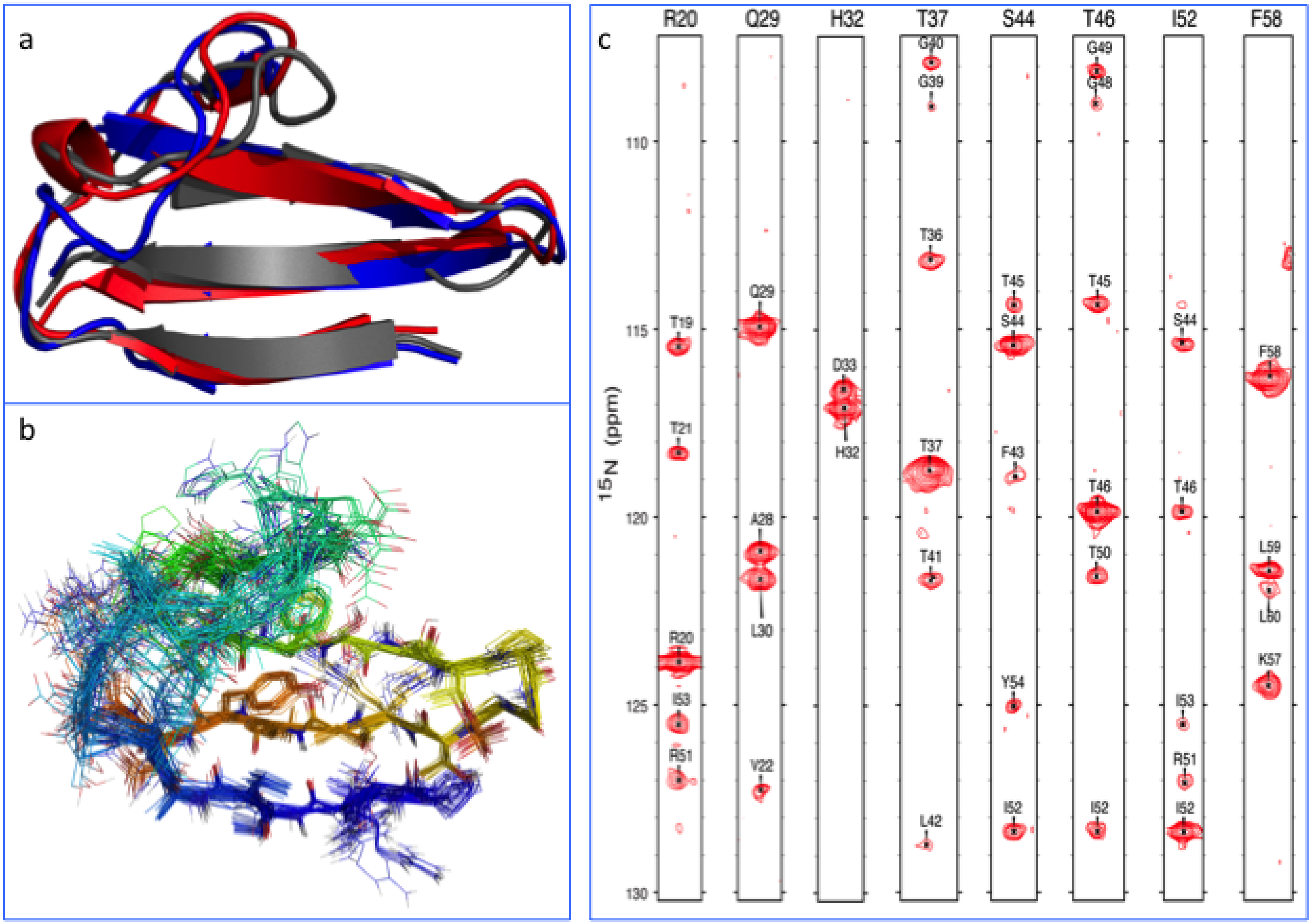
31. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
32. Goddard, T. D. & Kneller, D. G. SPARKY 3 (University of California, San Francisco, 2006).
33. Pascal, S. M., Muhandiram, D. R., Yamazaki, T., Formankay, J. D. & Kay, L. E. Simultaneous acquisition of 15N- and 13C-edited NOE spectra of proteins dissolved in H_2O. *J. Magn. Reson. B* **103,** 197–201 (1994).
34. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl Acad. Sci. USA* **105,** 4685–4690 (2008).
35. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44,** 213–223 (2009).
36. Kellogg, E. H., Leaver-Fay, A. & Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79,** 830–838 (2011).



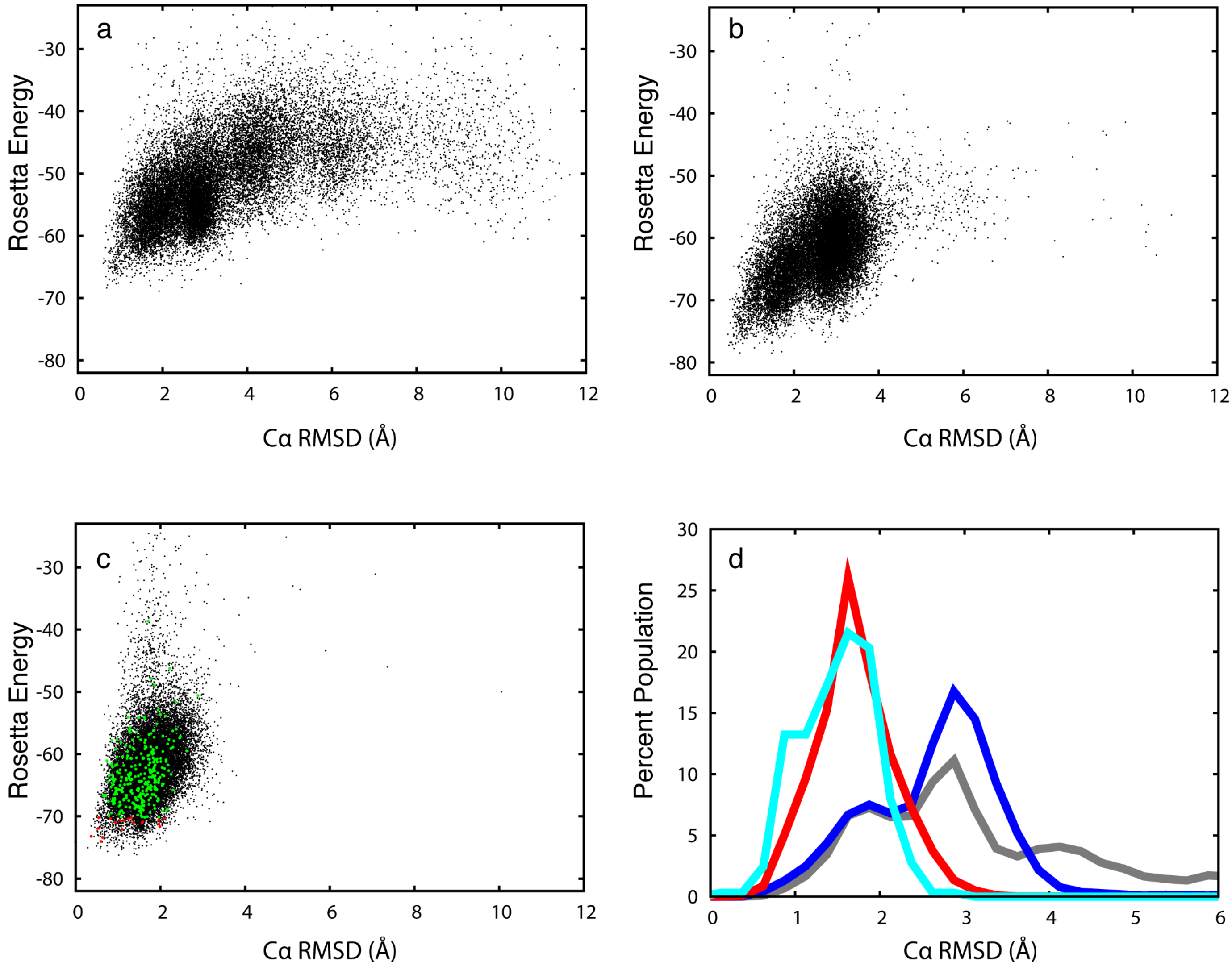
Extended Data Figure 1 | Effects of solution conditions on the structural and dynamic properties of phosphorylated 4E-BP2. **a**, **b**, Overlay of ^{1}H–^{15}N HSQC spectra of 4E-BP2 in 100 mM NaCl, 2 mM DTT, 1 mM EDTA, 1 mM benzamidine, 30 mM acetate pH 4 (red), 30 mM phosphate pH 6.2 (orange), 30 mM Tris pH 8.0 (green) or 30 mM CAPS pH 10.3 (blue) for (**a**) non-phosphorylated and (**b**) phosphorylated 4E-BP2. Because of the rapid ^{1}HN-solvent exchange of IDPs, ^{1}H-detected NMR experiments are usually performed at acidic pH. For non-phosphorylated 4E-BP2, all the peaks except one disappear by pH 10.3, while for phosphorylated 4E-BP2, most of the resonances for residues involved in hydrogen bonds upon folding remain visible. **c**, Effects of temperature on the structural and dynamic effects of phosphorylated 4E-BP2. Conformational exchange including *cis–trans* isomerization results in major and minor states of phosphorylated 4E-BP2. In addition to the assigned major states of both the folded and disordered regions of phosphorylated 4E-BP2, there remain many unassigned low-intensity peaks with ^{1}HN chemical shifts between 7.8 and 8.8 p.p.m., indicating that minor states contain significant disorder. The number of minor peaks decreases with increasing temperature, as shown in the ^{1}H–^{15}N HSQC spectra at 5 °C (black) and 35 °C (red), respectively. **d**, ^{1}HN temperature coefficients indicate intramolecular hydrogen bonding for many residues between P18 and R62. Chemical shift changes of ^{1}H–^{15}N HSQC spectra of pWT from 5 to 15 °C at pH 6.8 were used to calculate the ^{1}HN temperature coefficients. A horizontal line is plotted at −4.6 p.p.b. K^{-1}, with values above this line indicative of intramolecular hydrogen bonding[29] and missing data points representing proline residues.



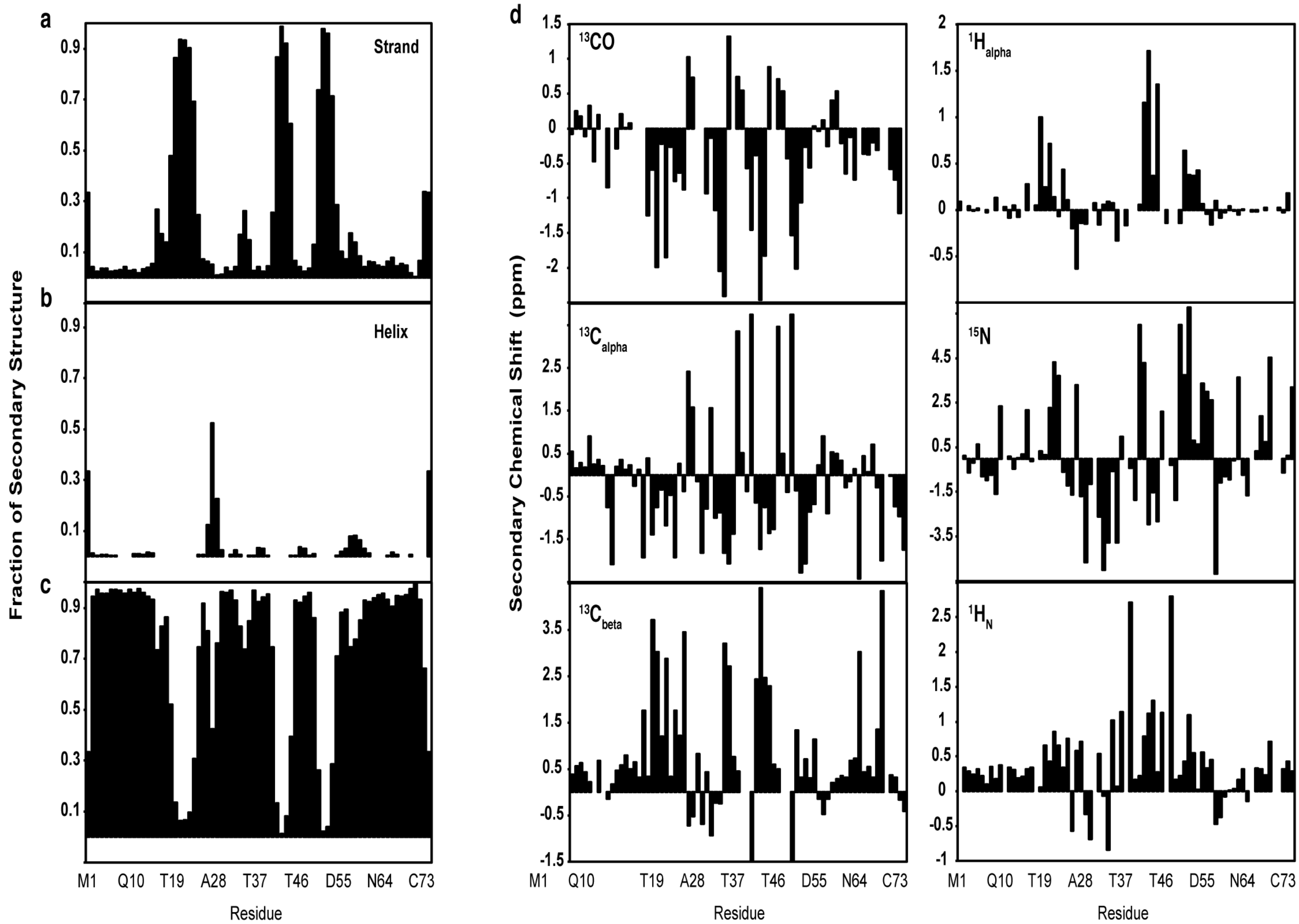
Extended Data Figure 2 | Phosphorylation, but not phosphomimetics, of both T37 and T46 are required to induce folding of 4E-BP2. a, **b**, Overlay of ^{1}H–^{15}N HSQC spectra of pWT with (**a**) pS65pT70pS83 and (**b**) pT37pT46. pS65pT70pS83 remains disordered as indicated by the lack of large ^{1}HN chemical shift dispersion of the folded state and absence of the downfield shifted G39 or G48 resonances. In contrast, pT37pT46 is nearly identical to pWT for residues P18–R62. **c**–**f**, Overlay of pT37pT46 with (**c**) pT37, (**d**) pT46, (**e**) T37D/T46D and (**f**) T37E/T46E. pT37 and pT46 are partly folded as indicated by the presence of only one of the downfield shifted G39 or G48 resonances in the insets, while aspartic acid or glutamic acid substitutions at T37 and T46 did not induce any folding, demonstrating that these are not good phosphomimetics. **g**, Binding of non-phosphorylated WT and pT37pT46 to unlabelled eIF4E. Overlay of two-dimensional ^{1}H–^{15}N HSQC spectra of non-phosphorylated WT (blue) and pT37pT46 (red) 4E-BP2 in complex with unlabelled eIF4E. Although both proteins are mostly disordered when bound to eIF4E, as evident from the poor ^{1}HN chemical shift dispersion, the pT37pT46 complex retains some structure, resulting in more chemical shift dispersion than WT.



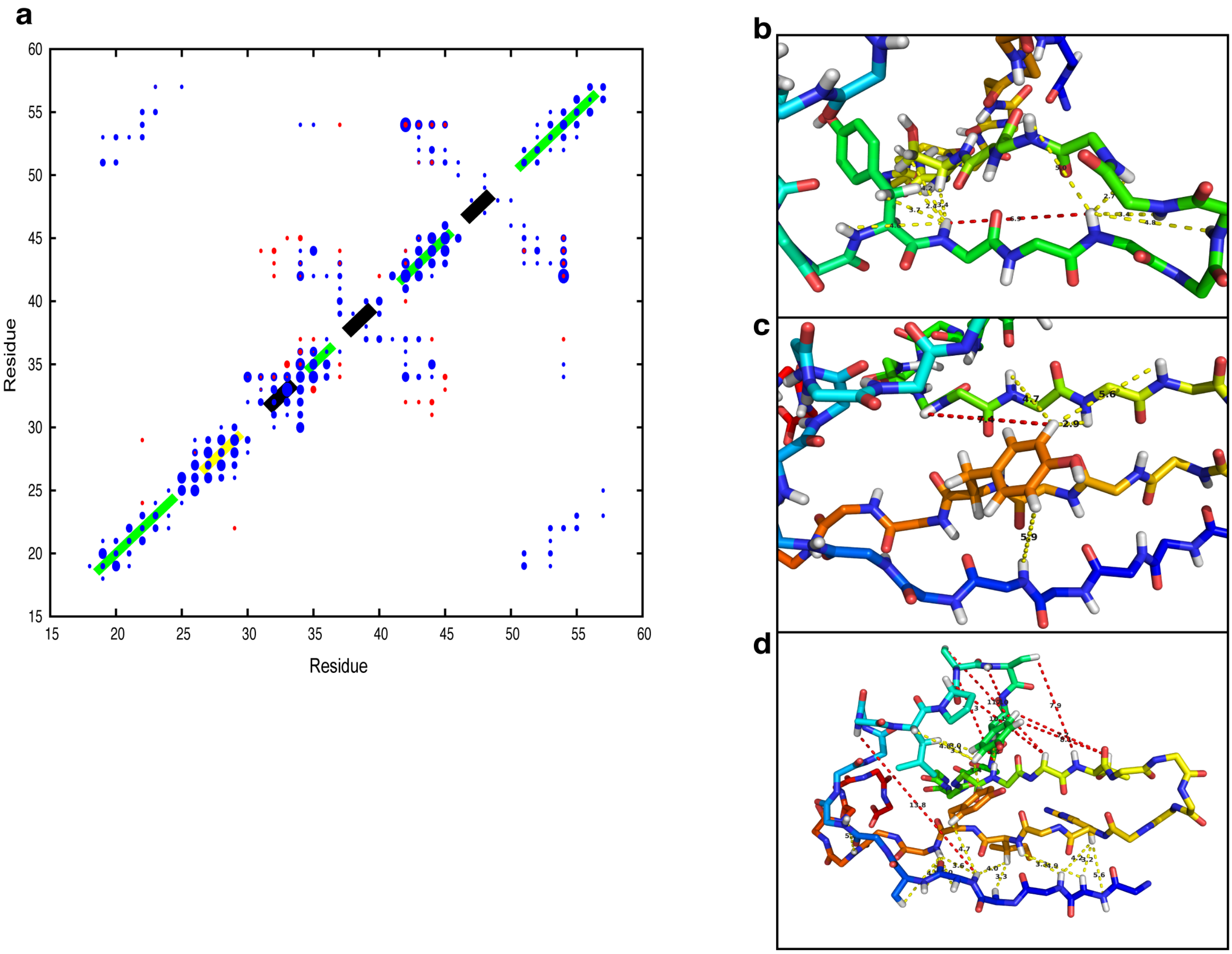
Extended Data Figure 3 | Structural models of phosphorylated 4E-BP2 calculated with CS-Rosetta and NMR data. **a**, Alignment of ribbon diagrams of the lowest energy structures from ~ 20,000 structural models using different inputs: ^{1}HN, ^{15}N, Cα and Cβ chemical shifts (grey), ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts (blue), and ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts as well as NOEs (red). **b**, Superposition of the final 20 lowest energy structures calculated using all chemical shifts and NOEs. Residues P18–K57 are shown using a rainbow colour spectrum from amino (N) to C termini. **c**, Examples of ^{1}HN–^{1}HN NOEs within the folded 4E-BP2. Shown are strips from ^{15}N-edited ^{15}N-NOESY demonstrating both short- and long-range interactions. Note that residues within the long loop connecting strands β1 and β2 contain no short- or long-range NOEs, indicating that it is very dynamic, consistent with low [^{1}H]–^{15}N NOEs. Not surprisingly, this loop shows the largest variation in the models (see **a**, **b**).



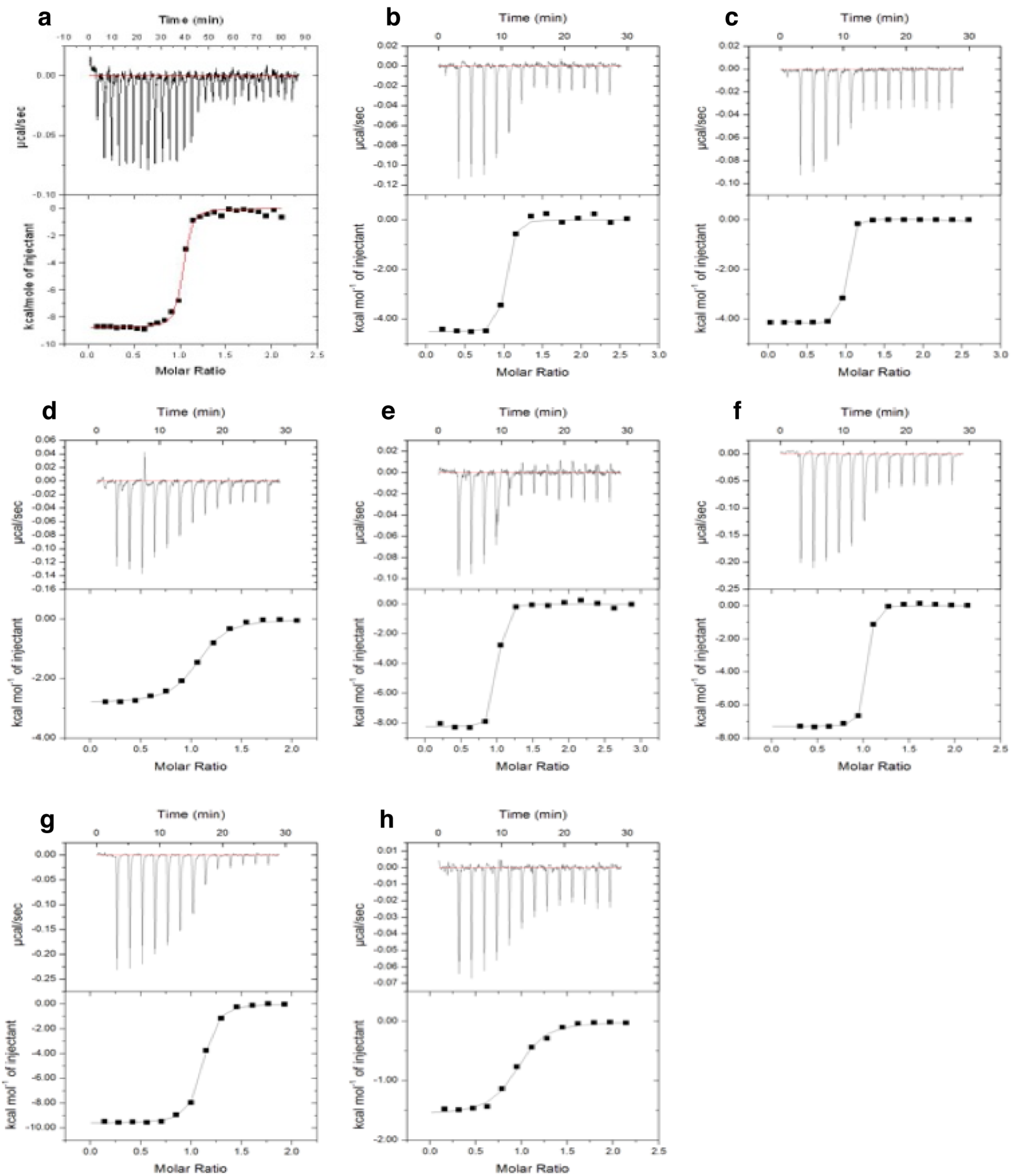
Extended Data Figure 4 | CS-Rosetta scores for calculations of the folded state of phosphorylated 4E-BP2 using different input data. CS-Rosetta2 was used to create models three separate times as new input data were acquired, each time with the observation that low Rosetta energy models converge to the same topology. Over these three sequential runs each addition of new data served primarily to drive sampling towards a previously observed energy minimum. **a**–**c**, CS-Rosetta energy for ~20,000 structural models as a function of Cα r.m.s.d. to the structure with the lowest energy point in the final ensemble using (**a**) ^{1}HN, ^{15}N, ^{13}Cα and ^{13}Cβ chemical shifts, (**b**) ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts, and (**c**) all chemical shifts and NOEs (final ensemble). In green are the best 5% of structures on the basis of agreement with NOEs, whereas in red are the 20 structures with the lowest NOE violations used to generate Extended Data Fig. 3b. Note that the CS-Rosetta energy plotted here is the empirical Rosetta energy value without chemical shift or NOE terms, reflecting the intrinsic energy rather than the fit to experimental data. **d**, Histograms showing the percentage distribution of structures with Cα r.m.s.d. as a function of Cα r.m.s.d. (going out to 6 Å) for the different CS-Rosetta input data: ^{1}HN, ^{15}N, ^{13}Cα and ^{13}Cβ chemical shifts (grey), ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts (blue), ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ, and ^{13}CO chemical shifts and NOEs without filtering (red) and with the final set of filters (cyan) for (1) the hydrogen bonding observed from the temperature coefficient measurements and (2) the best 5% by number of NOE satisfied.



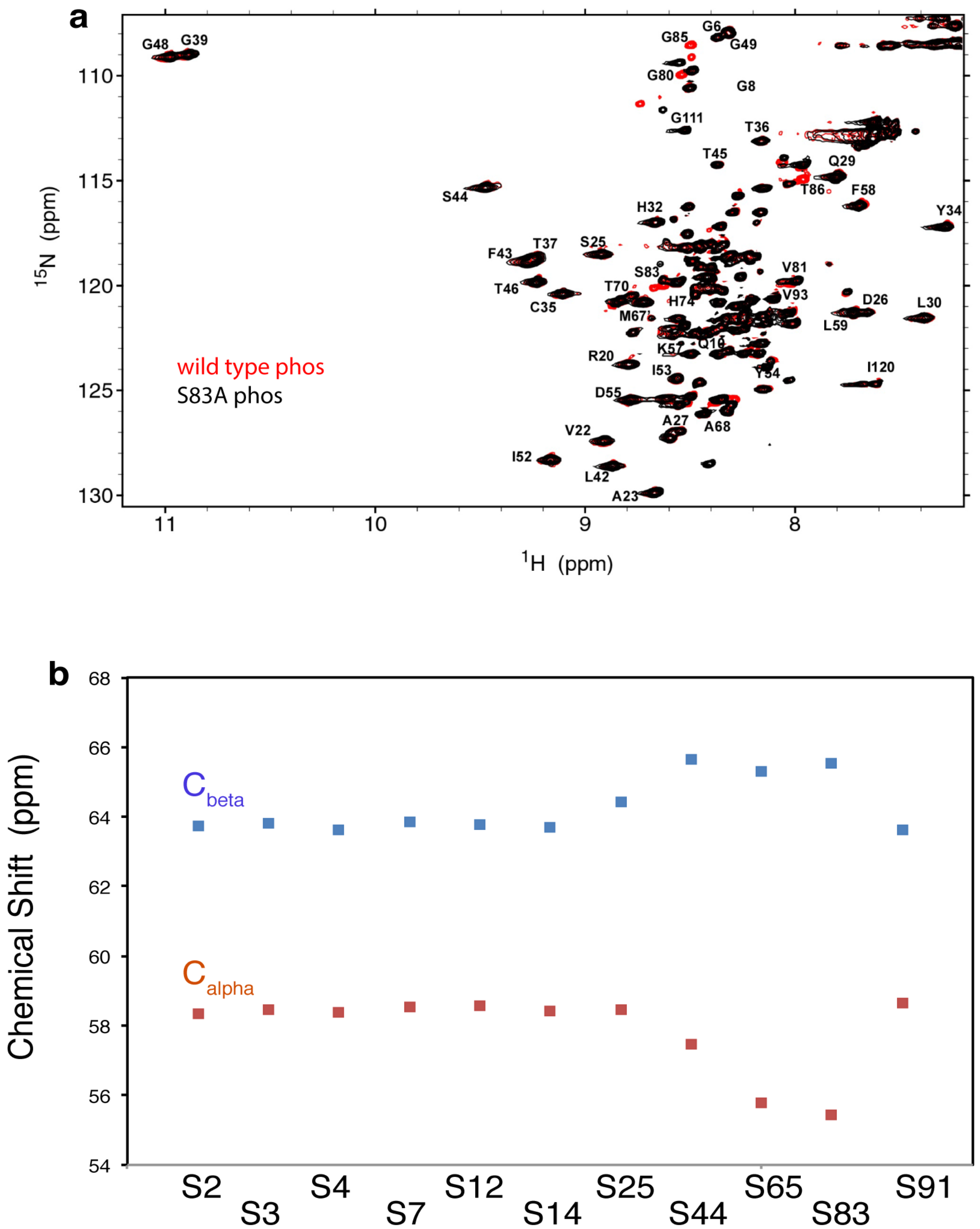
Extended Data Figure 5 | Chemical shifts and calculated secondary structure of phosphorylated 4E-BP2. Secondary chemical shifts define the topology of phosphorylated 4E-BP2 and validate the CS-Rosetta approach for structure determination. **a**–**c**, Fractional secondary structure as calculated by Talos+ as a function of residues number for (**a**) strand, (**b**) helix and (**c**) loop for residues 1–75 of phosphorylated 4E-BP2. **d**, Secondary chemical shifts for the folded region of phosphorylated 4E-BP2 from Talos+ as a function of residue number for ^{13}CO, ^{1}Hα, ^{13}Cα, ^{15}N, ^{13}Cβ and ^{1}HN shifts.



Extended Data Figure 6 | NOE violations for calculated structure of phosphorylated 4E-BP2. **a**, Contact map showing observed NOEs for each pair of residues for satisfied NOE restraints in blue and unsatisfied NOE restraints in red, with the areas of the circles proportional to the total number of NOEs in each case. Violations were calculated using distance boundaries of 5/6/7Å for strong, medium and weak NOEs, and the number of violations for each residue pair was either averaged across the ensemble (above the $x = y$ line) or by only counting restraints that were never satisfied in any of the models in the ensemble (below the $x = y$ line). The secondary structure of the protein is represented on the diagonal in green (β-strand), yellow (3_{10} helix) and black (turn) bars. **b**–**d**, Examples of NOE violations consistent with dynamic conformational exchange. A detailed look at individual NOE pairs (satisfied shown in yellow dashed lines, unsatisfied in red dashed lines) supports the conclusion that minor conformations contribute to the high number of violations, as consistently violated restraints conflict with the majority of the data that define the major conformation. For more information about the NOE violations and conformational exchange within phosphorylated 4E-BP2, see Supplementary Information.



Extended Data Figure 7 | ITC binding profiles of several 4E-BP2 constructs to eIF4E at 20 °C. **a**, WT; **b**, pT37; **c**, pT46; **d**, pT37pT46; **e**, T37D/T46D; **f**, T37E/T46E; **g**, pG39VG48V); **h**, pWT using competition with pS65pT70pS83.



Extended Data Figure 8 | Resonance assignments of phosphorylated residues. **a**, Overlay of ^{1}H–^{15}N HSQC spectra of pWT in red with phosphorylated S83A (pT37pT46pS65pT70) in black, showing the absence of the pS83 peak in pWT and other local changes. The blue arrow indicates the position of A83. **b**, Serine ^{13}Cα (red) and ^{13}Cβ (blue) chemical shifts in phosphorylated 4E-BP2. Although they did not show significant downfield shifts in the ^{1}HN–^{15}N HSQC spectrum (Fig. 1), S65 and S83 showed significant deviations from the random coil values compared with the other serines, consistent with phosphorylation. S25 and S44 also showed deviations as a result of the interactions within the folded domain (close in space to other phosphates), but not to the degree expected for a phosphorylated serine.

Extended Data Table 1 | Structural and energetic properties of phosphorylated 4E-BP2

(a) Percentage of satisfied NOE restraints for various distance boundaries in CS-Rosetta2 calculations.

Distance Boundaries (Å) (Strong/Medium/Weak)	Initial CSRosetta2 Ensemble (% restraints below boundaries)		Constrained CSRosetta2 Ensemble (% restraints below boundaries)	
	All NOEs	Long Range*	All NOEs	Long Range*
3.0/4.0/5.0	58.9	41.6	62.2	52.3
3.5/4.5/5.5	72.3	51.6	78.3	68.1
4.0/5.0/6.0	79.8	61.0	86.3	77.7
4.5/5.5/6.5	86.1	68.7	90.6	81.2
5.0/6.0/7.0	88.8	72.4	92.8	82.7
5.5/6.5/7.5	91.9	78.2	94.7	86.9
6.0/7.0/8.0	93.4	81.7	96.0	88.8

(b) Summary of NOE restraints and structural statistics

Number of unambiguous NOEs	450
Intraresidue ($\lvert i - j \rvert = 0$)	116
Short range ($\lvert i - j \rvert = 1$)	172
Medium-range ($2 \leqslant \lvert i - j \rvert \leqslant 9$)	95
Long-range ($\lvert i - j \rvert > 9$)	67
Number of ambiguous NOEs	6
Ramachandran plot statistics (%)	
Residues in most favored regions	93.5
Residues in additionally allowed regions	6.5
Residues in generously allowed regions	0
Residues in disallowed regions	0
Average RMSD to mean (Å)	
Backbone (residues P18-K57)	0.92 ± 0.24
Heavy atoms (residues P18-K57)	1.47 ± 0.22

(c) ΔΔG predictions of energetic effects of mutation of the TTPGGT motifs and canonical binding site using the initial CS-Rosetta structural model.

Amino Acid	Residue #	Mutate to / Predicted ΔΔG							
T	36 / 45	A	+1.1 / +1.3	D	+1.8 / +5.5	E	+0.66 / +4.8	S	+1.3 / +1.6
T	37 / 46	A	+1.5 /+0.62	D	+0.61/+0.78	E	+1.2 / +0.94	S	+1.4 / +0.77
P	38 / 47	A	+1.8 / +2.1	G	+3.1 / +3.6				
G	39 / 48	A	+6.8 / +3.4	V	+8.3 / +5.6	P	+15 / +120		
G	40 / 49	A	+6.4 /+0.29	V	+9.6 / -0.10	P	+22 / +13		
T	41 / 50	A	-0.42/+2.11	D	+0.64 / +3.5	E	+0.05 / +1.4	S	+0.39 / +2.7
Y	54	A	+5.0	F	+0.75	E	+1.5		

Extended Data Table 2 | ITC binding parameters for 4E-BP2 constructs to eIF4E

	ΔH (kcal/mol)	TΔS (kcal/mol)	K_d (nM)	ΔG (Binding & Unfolding)[a] (kcal/mol)	Folded?	Estimated Unfolding Free Energy[a] (kcal/mol)
WT	-8.81 ± 0.08	2.61	3.20 ± 0.6	-11.42	No	-
pT37	-4.51 ± 0.08	6.42	7.25 ± 1.7	-10.93	Partially	0.49
pT46	-4.14 ± 0.02	7.42	2.43 ± 0.5	-11.56	Partially	~ 0
pT37pT46	-2.83 ± 0.03	5.98	267 ± 32	-8.81	Yes	2.61[b]
T37D/T46D	-8.25 ± 0.09	3.08	3.89 ± 1.1	-11.33	No	-
T37E/T46E	-7.29 ± 0.04	3.93	4.37 ± 0.8	-11.22	No	-
pS65pT70pS83	-4.27 ± 0.04	6.39	11.3 ± 2.9	-10.66	No	-
p(G39V/G48V)	-9.62 ± 0.05	0.361	36.1 ± 3.5	- 9.98	No	-
pWT	-	-	12,320 ± 200	-6.59	Yes	4.83[b]

Enthalpy changes (ΔH), entropy changes (ΔS) and dissociation constants (K_d) from ITC measurements as well as combined binding and unfolding free energies derived from K_d values using $\Delta G = -RT\ln(1/K_d)$, with R the ideal gas constant and T the temperature, and estimated unfolding free energies using WT as a reference for pT37, pT46, pT37pT46 and pWT.

[a] ΔG values, which may include contributions from binding, unfolding of the β-sheet structure encompassing residues P18–R62 and folding of the helix encompassing the YXXXXLΦ sequence, were calculated from $\Delta G = \Delta H - T\Delta S = -RT\ln(1/K_d)$. The estimated unfolding free energy was obtained using $\Delta G - \Delta G_{WT}$, as for WT there was no contribution from unfolding.

[b] The enhanced stability of the fivefold phosphorylated (pWT) over pT37pT46, 4.8 versus 2.6 kcal mol^{-1} estimated unfolding free energy, is possibly due to long-range transient electrostatic interactions between the phosphorylated acidic C terminus with the basic folded domain. The predicted isoelectric point (pI) of non-phosphorylated full-length 4E-BP2 is 6.16; however, the region involved in the phosphorylation-induced folding is very basic (pI = 9.77), while the C terminus (S65–I120) is acidic (pI = 4.68). Phosphorylation is expected to drop the pI even further, potentially causing stabilizing transient electrostatic interactions between the folded region and the phosphorylated C terminus.

doi:10.1038/nature14219

Initiation of translation in bacteria by a structured eukaryotic IRES RNA

Timothy M. Colussi[1,2]†*, David A. Costantino[1,2]*, Jianyu Zhu[3]†, John Paul Donohue[3], Andrei A. Korostelev[3]†, Zane A. Jaafar[1], Terra-Dawn M. Plank[1]†, Harry F. Noller[3] & Jeffrey S. Kieft[1,2]

The central dogma of gene expression (DNA to RNA to protein) is universal, but in different domains of life there are fundamental mechanistic differences within this pathway. For example, the canonical molecular signals used to initiate protein synthesis in bacteria and eukaryotes are mutually exclusive[1]. However, the core structures and conformational dynamics of ribosomes that are responsible for the translation steps that take place after initiation are ancient and conserved across the domains of life[2]. We wanted to explore whether an undiscovered RNA-based signal might be able to use these conserved features, bypassing mechanisms specific to each domain of life, and initiate protein synthesis in both bacteria and eukaryotes. Although structured internal ribosome entry site (IRES) RNAs can manipulate ribosomes to initiate translation in eukaryotic cells, an analogous RNA structure-based mechanism has not been observed in bacteria. Here we report our discovery that a eukaryotic viral IRES can initiate translation in live bacteria. We solved the crystal structure of this IRES bound to a bacterial ribosome to 3.8 Å resolution, revealing that despite differences between bacterial and eukaryotic ribosomes this IRES binds directly to both and occupies the space normally used by transfer RNAs. Initiation in both bacteria and eukaryotes depends on the structure of the IRES RNA, but in bacteria this RNA uses a different mechanism that includes a form of ribosome repositioning after initial recruitment. This IRES RNA bridges billions of years of evolutionary divergence and provides an example of an RNA structure-based translation initiation signal capable of operating in two domains of life.

Bacteria cannot recognize the 'cap' on the 5′ end of eukaryotic messenger RNAs and eukaryotic ribosomes cannot use the Shine–Dalgarno sequence (SDS)[1] (Extended Data Fig. 1a). Although non-canonical mechanisms exist[3,4], there is no known translation initiation signal that can operate in multiple domains of life at any location in an mRNA. Despite this divergence there is strong conservation in the functional core of the ribosome, where mRNA and tRNAs interact and move[2]. In fact, the tRNAs used in elongation from bacteria and eukaryotes are interchangeable[5]. Therefore, we asked whether a structured RNA embedded in an mRNA sequence could interact with conserved ribosome features in the decoding groove and initiate translation in both bacteria and eukaryotes. Candidates for such RNAs are the intergenic region (IGR) IRESs from *Dicistroviridae* viruses. In eukaryotes, these IRESs act independently of a 5′ cap[6], adopt a functionally essential compact fold that docks within the ribosome[7–9] without initiation factors or a start codon[10–16], and partially mimic tRNA (Extended Data Fig. 1b, c)[12,17–19]. It is proposed that they drive translation initiation by co-opting the ribosome's conserved elongation cycle[17,19–22], and they operate in diverse eukaryotic systems[6,23].

We generated an inducible expression vector encoding a single mRNA containing two independent luciferase (LUC) reporters (Extended Data Fig. 1d)[24], and verified that it allowed the simultaneous measurement of initial rates of production of each protein (Extended Data Figs 2 and 3). We used this construct to test whether an IGR IRES RNA can drive translation in live bacteria. *Renilla* luciferase (RLUC) was placed to initiate translation from an SDS (and 'enhancer' sequence), and firefly luciferase (FLUC) was placed after a wild-type *Plautia stali* intestine virus (PSIV) IGR IRES. There was some production of both LUCs before induction (due to expected 'leaky expression'; Extended Data Fig. 4), but induction resulted in a marked increase in both reporters; the production

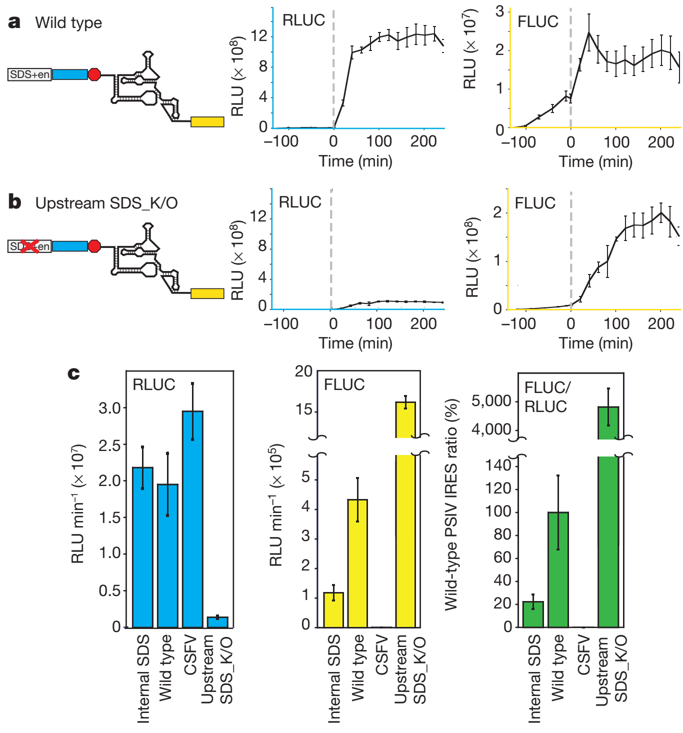


Figure 1 | Translation initiation assays in bacteria. a, Full-length wild-type IRES. Left, diagram of the construct. en, enhancer. Middle, graph shows relative light units (RLU) from the upstream RLUC as a function of time. Dashed grey line is $t = 0$, the point of induction. The trace is the average signal of at least three experiments, with error bars showing 1 standard deviation (s.d.) from the mean. Right, graph shows FLUC expression from the IRES. **b**, Diagram and traces from the Upstream SDS_K/O mutant. Note the change in scale of the y-axis for FLUC. **c**, Initial rates of RLUC and FLUC production, and the FLUC/RLUC ratio for the indicated constructs. Error bars represent 1 s.d. from the mean from three biological replicates. See Extended Data Figs 2 and 5 for diagrams and raw traces of the Internal SDS and CSFV constructs.

[1]Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, Colorado 80045, USA. [2]Howard Hughes Medical Institute, University of Colorado Denver School of Medicine, Aurora, Colorado 80045, USA. [3]Center for Molecular Biology of RNA and Department of Molecular, Cell and Developmental Biology, Sinsheimer Labs, University of California at Santa Cruz, Santa Cruz, California 95064, USA. †Present addresses: Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115, USA (T.M.C.); Cocrystal Discovery, Inc., Mountain View, California 94043, USA (J.Z.); RNA Therapeutics Institute, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA (A.A.K.); Department of Reproductive Medicine, University of California at San Diego, La Jolla California 92093, USA (T.-D.M.P.).
*These authors contributed equally to this work.

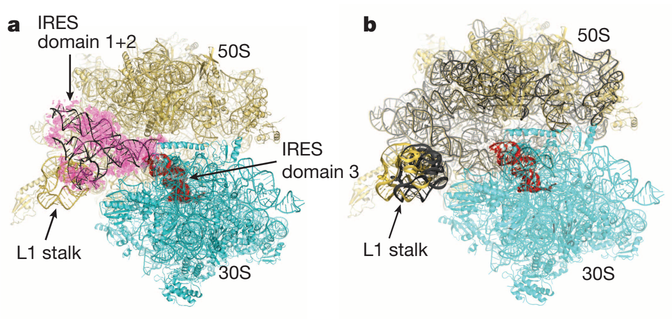


Figure 2 | IRES•70S ribosome structure. **a**, Crystal structure of a full-length PSIV IGR IRES bound to *Thermus thermophilus* 70S ribosomes. Cyan, small subunit; yellow, large subunit; red, PSIV IRES domain 3; grey, density corresponding to domain 3; magenta, unbiased difference $F_o - F_c$ density corresponding to domain 1+2, with the crystal structure of PSIV IGR IRES domain 1+2 (black) docked as a rigid body[26]. **b**, Superimposition of crystal structures of the PSIV IGR IRES•70S ribosome complex (this work) and the 70S ribosome. Yellow, IRES-bound 50S subunit. Domain 1+2 shifts the L1 stalk relative to its position in tRNA-bound complexes by ~15 Å.

of FLUC is consistent with translation beginning at the IRES (Fig. 1c and Extended Data Fig. 2). Removing the RLUC-driving SDS (Upstream SDS_K/O; all mutants shown in Extended Data Fig. 5) diminished production of RLUC, but FLUC production increased more than tenfold (Fig. 1b; all raw LUC data in Extended Data Table 1a), which we attributed to decreased competition for ribosomes and to ribosomes initiating independently at the IRES. Replacing the IGR IRES with the IRES from classical swine fever virus (CSFV) resulted in negligible FLUC production (Extended Data Fig. 2), demonstrating specificity for the IGR IRES.

A source of initiation from the IGR IRES could be a 'cryptic' SDS in the purine-rich sequence between the IRES and the FLUC start codon (Extended Data Fig. 6). FLUC production from this SDS-like sequence alone was at ~30% of the wild-type IRES, not enough to account for all FLUC produced from the IRES. Mutating this SDS-like sequence in the context of the full IRES decreased FLUC production, but translation was still higher than from an SDS or the SDS-like sequence. Thus, the structured IRES can drive FLUC production without the SDS-like sequence, but both probably contribute to function when present.

To determine the structural basis for IGR IRES activity in bacteria, we solved the crystal structure of the full-length IRES RNA•70S ribosome complex to 3.8 Å resolution. In eukaryotes, IGR IRES domains 1 and 2 (domain 1+2) contact both subunits, whereas domain 3 mimics an mRNA–tRNA interaction on the small subunit (Extended Data Fig. 1b)[7,8,10,11,19,25]. We observed electron density for domain 3 in the P site as in the crystal structure of isolated domain 3 bound to 70S ribosomes[19] (Fig. 2a and Extended Data Fig. 7); this may represent an initiation-state or translocated IRES. The density of domain 1+2 was weak but its location could be modelled using the crystal structure of unbound PSIV IGR IRES domain 1+2 (Fig. 2a)[26]. The location of domain 1+2 in the 70S ribosome differs from IGR IRES•80S ribosome complexes, with domain 3 in the A site[22,27]. In 80S ribosomes, domain 1+2 interacts with the eukaryotic-specific ribosomal protein eS25 and the L1 stalk[10,11,28,29], which is structurally distinct from that in bacterial ribosomes[30]. In the full-length IRES•70S structure, the L1 stalk is displaced ~15 Å compared with the structure containing domain 3 only (Fig. 2b). The absence of eS25 and differences in the L1 stalk may be responsible for the partial disorder and location of the IRES. Nonetheless, the structure clearly illustrates that the compactly folded IRES can bind in the tRNA-binding sites of bacterial ribosomes.

The compact structure of the IGR IRES is essential for its function in eukaryotes[25,26], and the IRES•70S structure suggested that this is also true in bacteria. To test this, we disrupted two pseudoknots essential for the compact structure of the IRES, both individually (PK1_K/O, PK2_K/O) and together (PK1+PK2_K/O), and measured activity (Fig. 3a, b and Extended Data Fig. 8a)[10]. FLUC production decreased in all three mutants, with FLUC production in the double mutant at a level that could be accounted for by activity from the cryptic SDS-like sequence. Indeed, disruption of both pseudoknots and the SDS-like sequence (Downstream SDS-like_K/O+PK1+PK2_K/O) abrogated IRES activity (Extended Data Fig. 6). Isolated IRES domain 3 operated similarly to the domain-1+2-disrupting mutant (PK2_K/O). Thus, IGR IRES translation in bacteria depends on a compact RNA structure and although domain 1+2 is poorly ordered in the crystal, it may be required to form transient interactions with the ribosome.

We explored the putatively transient IGR IRES•70S interactions using translationally competent cell-free extracts. In rabbit reticulocyte lysate (RRL; positive control) the IRES forms 80S ribosomes both in the presence and absence of a non-hydrolysable GTP analogue (GMPPNP)

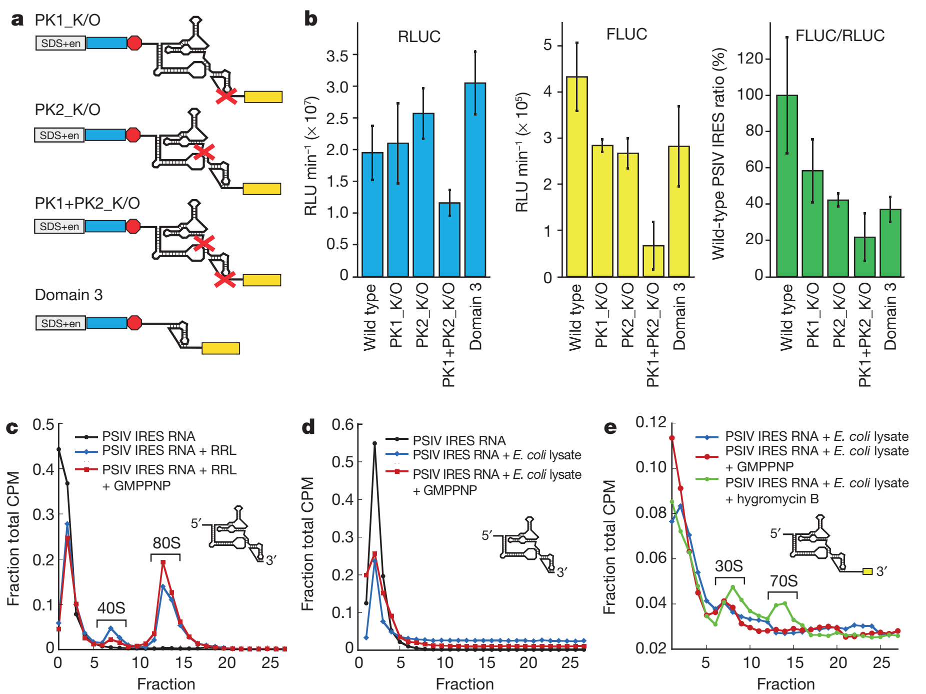


Figure 3 | Importance of IRES structure and ribosome binding. **a**, IRES constructs with structural domains disrupted or removed. en, enhancer. **b**, Rates of LUC production and LUC ratio. Error bars represent 1 s.d. from the mean from three biological replicates. **c**, Ribosome assembly assay with the PSIV IGR IRES in RRL, resolved on a sucrose gradient. Locations of complexes are indicated. CPM, counts per minute. **d**, As for **c**, but in *E. coli* lysate. **f**, As for **d**, but with an IRES RNA containing a downstream sequence to include the FLUC start codon. **c**–**e**, The addition of GMPPNP or hygromycin B is indicated. Data from one experiment are shown.

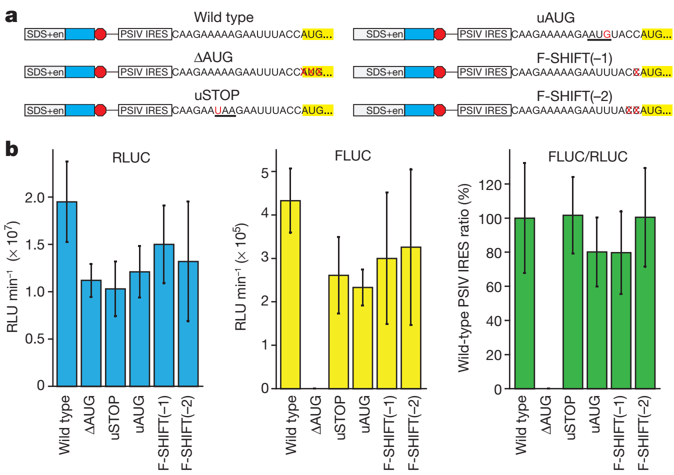


Figure 4 | Location of initiation on an IGR IRES in bacteria. **a**, Constructs designed to determine the location of initiation. For uAUG and uSTOP, the start and stop codons are underlined. en, enhancer. **b**, Rates of LUC production and LUC ratio from these constructs. Error bars represent 1 s.d. from the mean from three biological replicates.

(Fig. 3c). In contrast, 70S formation on the IRES in *Escherichia coli* lysate was virtually undetectable (Fig. 3d). We repeated the experiment with an IRES RNA containing the FLUC AUG and several codons downstream of the IRES to allow initiation to occur and stabilize the resultant complexes. Both IRES•70S complexes and IRES•30S complexes formed in the presence of the elongation inhibitor hygromycin B (Fig. 3e). In the *E. coli* lysate, the amount of IRES–ribosome complex is low compared to that observed for the RRL, consistent with formation of an unstable or transient complex.

In eukaryotes, IGR IRES-driven translation begins directly on the IRES and is proposed to co-opt the ribosome's elongation cycle[17,19,21,22]; we asked whether this is true in bacteria, in which the IRES–ribosome interactions appear different and transient. Removal of the FLUC start codon located 15 nucleotides downstream of the IRES structure (ΔAUG) resulted in a complete loss of FLUC production, while a stop codon placed upstream of the FLUC start codon (uSTOP) had little effect (Fig. 4a, b and Extended Data Fig. 8b). Removal of 1 or 2 nucleotides just upstream of the FLUC AUG (F-SHIFT(-1) and F-SHIFT(-2)) had little effect. These results are consistent with translation in bacteria beginning on the FLUC AUG, not directly at the IRES on a non-AUG codon. This implies a repositioning of the ribosome from the IRES to the FLUC start codon. To explore this, we created a construct with an out-of-frame start codon between the IRES and the start codon (uAUG); this mutation decreased activity but not to the degree that would be expected if this codon were being used efficiently. The source of this discrimination is not clear, but we posit that selection of the FLUC AUG is assisted by the nearly ideally positioned cryptic SDS-like sequence upstream. Constructs with alterations between the IRES and FLUC start codon all had decreased activity in the context of the PK1+PK2_K/O mutation (Extended Data Fig. 9), indicating that IRES structural integrity remains necessary for their function.

The mechanism of the IRES studied here in bacteria is more primitive than in eukaryotes. We propose that the structured IRES RNA forms interactions with bacterial ribosomes that are transient and weaker than the highly tuned interactions that occur in eukaryotes, but allow internal entry of the ribosome to the message. Recruited subunits or ribosomes are repositioned to a downstream start codon where protein synthesis starts. That a compact IRES RNA can use this primitive mechanism suggests that RNA structure-driven or structure-assisted initiation may be used in potentially all domains of life, driven by diverse RNAs perhaps possessing tRNA-like character or decoding groove binding capability.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 January 2014; accepted 7 January 2015.

Published online 4 February 2015.

1. Hershey, J. W., Sonenberg, N. & Mathews, M. B. Principles of translational control: an overview. *Cold Spring Harb. Perspect. Biol.* **4,** a011528 (2012).
2. Melnikov, S. *et al.* One core, two shells: bacterial and eukaryotic ribosomes. *Nature Struct. Mol. Biol.* **19,** 560–567 (2012).
3. Malys, N. & McCarthy, J. E. Translation initiation: variations in the mechanism can be anticipated. *Cell. Mol. Life Sci.* **68,** 991–1003 (2011).
4. Moll, I., Grill, S., Gualerzi, C. O. & Blasi, U. Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Mol. Micr. obiol.* **43,** 239–246 (2002).
5. Berthelot, F., Bogdanovsky, D., Schapira, G. & Gros, F. Interchangeability of factors and tRNA's in bacterial and eukaryotic translation initiation systems. *Mol. Cell. Biochem.* **1,** 63–72 (1973).
6. Hertz, M. I. & Thompson, S. R. Mechanism of translation initiation by *Dicistroviridae* IGR IRESs. *Virology* **411,** 355–361 (2011).
7. Schüler, M. *et al.* Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nature Struct. Mol. Biol.* **13,** 1092–1096 (2006).
8. Spahn, C. M. *et al.* Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. *Cell* **118,** 465–475 (2004).
9. Pestova, T. V., Lomakin, I. B. & Hellen, C. U. Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. *EMBO Rep.* **5,** 906–913 (2004).
10. Jan, E. & Sarnow, P. Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *J. Mol. Biol.* **324,** 889–902 (2002).
11. Nishiyama, T. *et al.* Structural elements in the internal ribosome entry site of *Plautia stali* intestine virus responsible for binding with ribosomes. *Nucleic Acids Res.* **31,** 2434–2442 (2003).
12. Jan, E. *et al.* Initiator Met-tRNA-independent translation mediated by an internal ribosome entry site element in cricket paralysis virus-like insect viruses. *Cold Spring Harb. Symp. Quant. Biol.* **66,** 285–292 (2001).
13. Thompson, S. R., Gulyas, K. D. & Sarnow, P. Internal initiation in *Saccharomyces cerevisiae* mediated by an initiator tRNA/eIF2-independent internal ribosome entry site element. *Proc. Natl Acad. Sci. USA* **98,** 12972–12977 (2001).
14. Wilson, J. E., Pestova, T. V., Hellen, C. U. & Sarnow, P. Initiation of protein synthesis from the A site of the ribosome. *Cell* **102,** 511–520 (2000).
15. Sasaki, J. & Nakashima, N. Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc. Natl Acad. Sci. USA* **97,** 1512–1515 (2000).
16. Sasaki, J. & Nakashima, N. Translation initiation at the CUU codon is mediated by the internal ribosome entry site of an insect picorna-like virus *in vitro*. *J. Virol.* **73,** 1219–1226 (1999).
17. Costantino, D. A., Pfingsten, J. S., Rambo, R. P. & Kieft, J. S. tRNA–mRNA mimicry drives translation initiation from a viral IRES. *Nature Struct. Mol. Biol.* **15,** 57–64 (2008).
18. Jan, E., Kinzy, T. G. & Sarnow, P. Divergent tRNA-like element supports initiation, elongation, and termination of protein biosynthesis. *Proc. Natl Acad. Sci. USA* **100,** 15410–15415 (2003).
19. Zhu, J. *et al.* Crystal structures of complexes containing domains from two viral internal ribosome entry site (IRES) RNAs bound to the 70S ribosome. *Proc. Natl Acad. Sci. USA* **108,** 1839–1844 (2011).
20. Hellen, C. U. IRES-induced conformational changes in the ribosome and the mechanism of translation initiation by internal ribosomal entry. *Biochim. Biophys. Acta* **1789,** 558–570 (2009).
21. Yamamoto, H., Nakashima, N., Ikeda, Y. & Uchiumi, T. Binding mode of the first aminoacyl-tRNA in translation initiation mediated by Plautia stali intestine virus internal ribosome entry site. *J. Biol. Chem.* **282,** 7770–7776 (2007).
22. Fernández, I. S., Bai, X. C., Murshudov, G., Scheres, S. H. & Ramakrishnan, V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **157,** 823–831 (2014).
23. Deniz, N., Lenarcic, E. M., Landry, D. M. & Thompson, S. R. Translation initiation factors are not required for *Dicistroviridae* IRES function *in vivo*. *RNA* **15,** 932–946 (2009).
24. Stoneley, M., Paulin, F. E., Le Quesne, J. P., Chappell, S. A. & Willis, A. E. c-Myc 5′ untranslated region contains an internal ribosome entry segment. *Oncogene* **16,** 423–428 (1998).
25. Costantino, D. & Kieft, J. S. A preformed compact ribosome-binding domain in the cricket paralysis-like virus IRES RNAs. *RNA* **11,** 332–343 (2005).
26. Pfingsten, J. S., Costantino, D. A. & Kieft, J. S. Structural basis for ribosome recruitment and manipulation by a viral IRES RNA. *Science* **314,** 1450–1454 (2006).
27. Koh, C. S., Brilot, A. F., Grigorieff, N. & Korostelev, A. A. Taura syndrome virus IRES initiates translation by binding its tRNA-mRNA-like structural element in the ribosomal decoding center. *Proc. Natl Acad. Sci. USA* **111,** 9139–9144 (2014).
28. Nishiyama, T., Yamamoto, H., Uchiumi, T. & Nakashima, N. Eukaryotic ribosomal protein RPS25 interacts with the conserved loop region in a dicistroviral intergenic internal ribosome entry site. *Nucleic Acids Res.* **35,** 1514–1521 (2007).
29. Landry, D. M., Hertz, M. I. & Thompson, S. R. RPS25 is essential for translation initiation by the *Dicistroviridae* and hepatitis C viral IRESs. *Genes Dev.* **23,** 2753–2764 (2009).
30. Ben-Shem, A. *et al.* The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334,** 1524–1529 (2011).

Acknowledgements We thank the members of the Kieft laboratory for insight and discussion and the staff at the Advanced Photon Source for their support. The original PSIV IGR IRES-containing plasmid was from N. Nakashima and the source of the luciferase genes was a plasmid from A. Willis. This work was supported by grants GM-17129 and GM-59140 from the National Institutes of Health (NIH) and MCB-723300 from the National Science Foundation (to H.F.N.), grant GM-103105 from the NIH (to A.A.K.), and grants GM-97333 and GM-81346 from the NIH (to J.S.K.). J.S.K. is an Early Career Scientist of the Howard Hughes Medical Institute. T.-D.M.P. was an American Heart Association Predoctoral Scholar (10PRE260143).

Author Contributions T.M.C. and J.S.K. designed the experiments and the constructs tested. T.M.C. and D.A.C. conducted the bacterial functional assays. Clones were generated by T.M.C., T.-D.M.P. and Z.A.J. J.S.K. performed the ribosome association assays. Ribosomes were purified, crystals grown, and the structure solved by J.P.D., J.Z. and A.A.K. under the supervision of H.F.N. J.S.K. provided overall supervision and guidance, and together with T.M.C. and D.A.C. wrote the manuscript with input from all authors.

Author Information Atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank under accession number 4XEJ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.S.K. (jeffrey.kieft@ucdenver.edu).

METHODS

Plasmid construction. DNA containing the *Plautia stali* intestine virus (PSIV) IGR IRES (nucleotides 6000–6195) between the RLUC and FLUC coding sequences was ligated into the KpnI and SacI sites of a pET30a vector (Novagen) using T4 DNA Ligase (New England Biolabs). The resultant construct contained 15 nucleotides of sequence between the 3′ end of the IRES (designated as the 3′ end of pseudoknot 1, nucleotide 6195) and the AUG start codon of the FLUC open reading frame (ORF).

Generation of mutants. Mutants were generated using several methods. First, PCR with appropriate forward and reverse primers (IDT) was used to generate two halves of the desired sequence. The halves were annealed and amplified by PCR using the T7 and T7 terminator sequencing primers. The resultant DNA was then ligated into the same pET30a vector using the above restriction sites.

Second, site-directed mutagenesis using the QuikChange (Agilent) mutagenesis strategy using appropriate primer pairs.

Third, insertion of PCR-amplified DNA or synthesized gBlock gene fragments (IDT) into the dual-LUC-containing pET30a vector between the SpeI and NcoI sites (between the *Renilla* and firefly genes) using a ligation-independent cloning (LIC) method, In-Fusion HD Cloning Plus (Clontech Laboratories). PCR products or gBlocks contained sequence overlapping 12 base pairs (bp) 5′ of the SpeI site and 12 bp 3′ of the NcoI site of the vector. Assembled constructs maintained both restriction sites.

Fourth, for the T7 knockout construct, a pET30a vector containing a mutated T7 promoter (TAAATGGTGTCTGAATTC) was synthesized (DNA 2.0) and DNA coding for the wild-type PSIV flanked by the two LUC genes was amplified by PCR. The PCR product was inserted between the KpnI and SacI sites in the mutated T7 vector by LIC.

Fifth, the mutant in which the PSIV IGR IRES was replaced by a SDS (without enhancer sequence) was generated by ligating the DNA fragment into the pET30a/dual-LUC vector using the SpeI and NcoI sites.

Bacterial cell culture. Rosetta DE3 cells (Novagen) were transformed with the plasmids described earlier and grown overnight in 5 ml Luria broth (LB) with kanamycin (Fisher) at 37 °C with constant agitation to generate a starter culture. To start the experiment, 50 ml of LB containing kanamycin was inoculated with 1 ml of the overnight starter culture. The 50 ml cultures were grown with agitation at 37 °C to an absorbance at 600 nm of 0.6 (measured on a Thermo Scientific NanoDrop 2000c spectrophotometer). The cultures were induced with 1 mM isopropyl-β-D-thiogalactoside (IPTG) (Gold Bio) and allowed to grow for 4 h. Samples (50 μl) were taken at 10–30 min intervals.

Measurement of LUC activity. At each time point, 50 μl of cell culture was removed, the cells were pelleted by centrifugation, and the supernatant was removed. Cells were resuspended in 300 μl 1× Passive Lysis Buffer (PLB, Promega). Twenty microlitres of the resultant cell lysate was added to a 96-well microplate (Greiner Bio-One). The dual-LUC assay was performed by first adding 100 μl LAR II (Promega) to measure FLUC activity, then 100 μl of Stop & Glo reagent (Promega) was added to measure RLUC activity. The assay was performed and measurements were taken using a Promega Glomax Multi+ detection system.

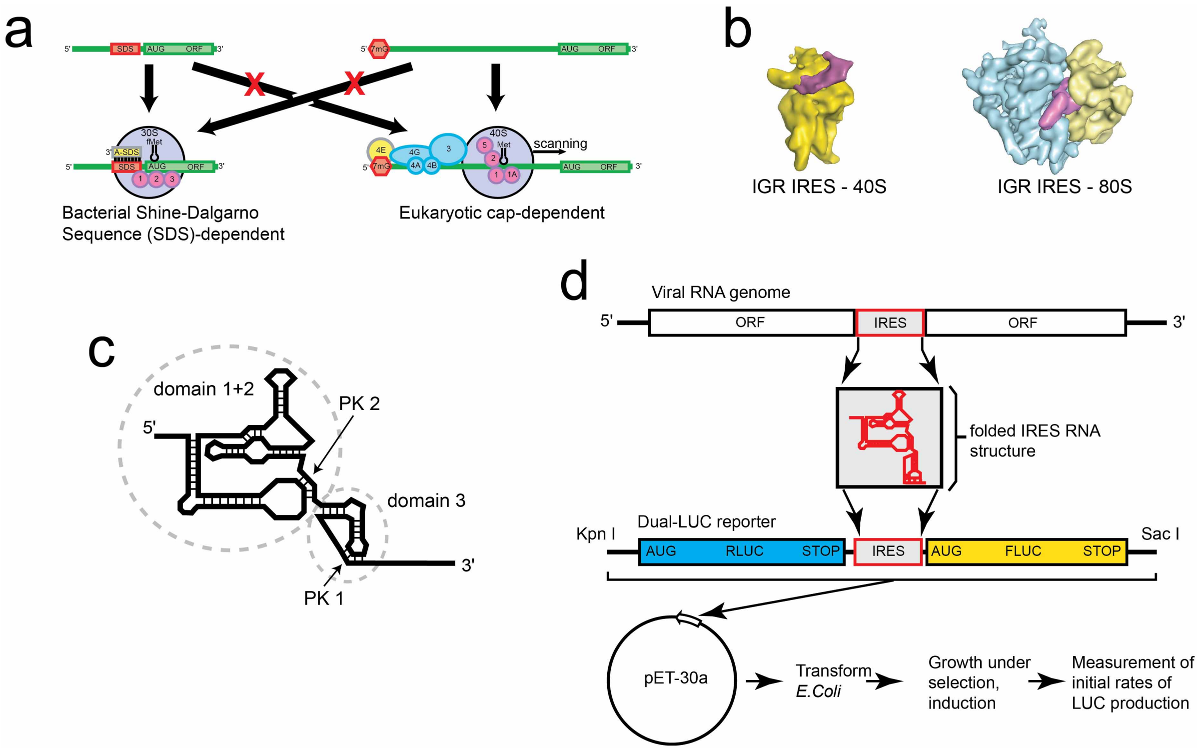
Determination of IRES activity. FLUC and RLUC activity (expressed as RLU) were graphed as a function of time for each culture using the program KaleidaGraph. The initial rate of FLUC and RLUC production was determined using the data from the first 30–40 min after induction. LUC production was generally linear over this time scale after a 5–10 min lag. IRES activity was then calculated as the ratio of the initial rate of FLUC to RLUC for each culture. Ratios from individual independent cultures were averaged. Bar graphs represent averages from at least three independent cultures; error bars depict one standard deviation from the mean. This method corrects for variation in growth, induction, and potential protein stability differences between cultures.

RNA transcription and purification for ribosome assembly assay. DNA templates for *in vitro* transcription were generated by PCR using a plasmid containing the wild-type PSIV IRES as the template and primers designed to amplify just the DNA of interest under the control of a T7 RNA polymerase promoter. The resultant PCR-generated DNA template was used in *in vitro* transcription reactions. RNA was purified from raw transcription reactions by high-performance liquid chromatography (HPLC). The first RNA used in assembly assays contained nucleotides 6000–6195 of the PSIV IGR IRES, and the second contained this same sequence, plus the sequence GAAAAAGAATTTACCATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAAGATGGAACCGCTGGAGAGC downstream of the IRES.

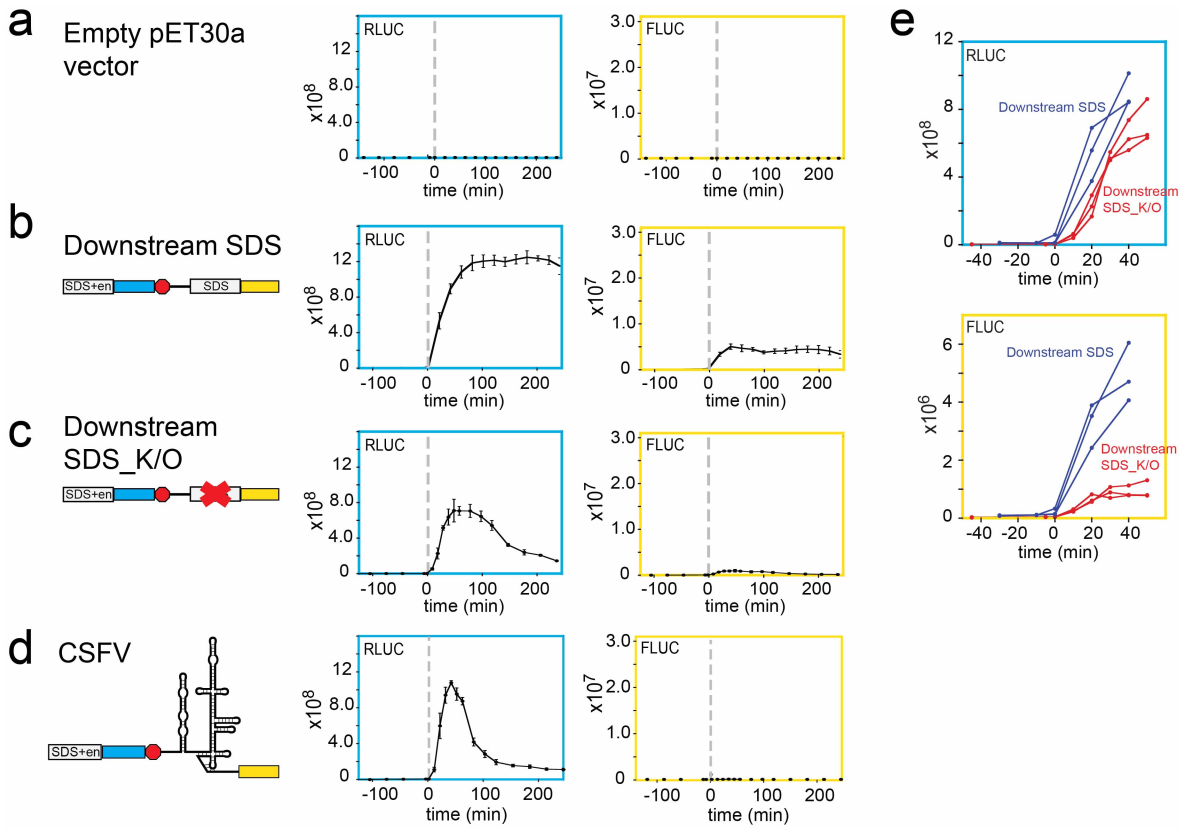
Ribosome assembly assay. RNA for use in assembly assays was 5′-end radiolabelled with P-32 using T4 polynucleotide kinase (NEB), purified by gel electrophoresis, and diluted to 1,000 CPM μl^{-1}. For the assays in RRL, 1 μl radiolabelled RNA was combined with 30 μl of lysate supplemented with amino acids. For the reaction with GMPPNP, 5 μl of a 20 mM stock of the analogue was added to achieve a final concentration of 2 mM, and an equimolar amount of $MgCl_2$ was added. For the reaction with hygromycin B, 2 μl of a 50 mg ml^{-1} stock was added to a final concentration of 2 mg ml^{-1}. RNase-free water was added to a total final volume of 50 μl. For the reactions in *E. coli* lysate, Promega product #L1030 was used. One microlitre of labelled RNA was added to 15 μl of lysate and 20 μl of S30 premix supplemented with 5 μl of the amino acid mix and 1 μl of RNasin RNase inhibitor (Promega). For the reactions with GMPPNP or antibiotic, the same amounts were added as for the RRL reactions. Reactions were incubated at 30 °C for 5 min, then 250 μl of ice-cold dilution buffer (40 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT)) was added and the reactions were immediately loaded on 15–30% sucrose gradients in dilution buffer. Gradients were centrifuged at 35,000 r.p.m. for 4 h in an SW41 rotor, then fractionated using a BioComp system. The amount of radiation in each fraction was measured and used to generate the plots. According to the manufacturer, this lysate contains substantial RNase activity; we attempted to mitigate this effect using RNase inhibitors. However, we were unable to fully eliminate the activity.

Crystallographic data collection and structure determination. 70S ribosomes were purified and the 70S•PSIV IRES complex was prepared and crystallized essentially as previously described[19]. The IRES RNA used contained nucleotides 6000–6195 of the PSIV viral RNA. X-ray diffraction data were collected at beamline 23 ID-B at the Advanced Photon Source at Argonne National Laboratory, using an X-ray wavelength of 1.033 Å and an oscillation angle of 0.2°. For determining the structure of the 70S•PSIV IRES complex, one data set obtained from a single crystal was integrated and scaled using XDS[31]. 0.4 per cent of the reflections were marked as test-set (R_{free} set) reflections and used for cross-validation throughout refinement. The previously determined X-ray structure of the 70S ribosome bound with domain 3 of the PSIV IRES, obtained from the same crystal form, was used as a molecular replacement model[19]. Domain 3 of the IRES and L1 stalk were removed from this starting model. Initial $F_o - F_c$ difference maps were calculated after rigid-body and simulated-annealing refinement was performed using two-fold non-crystallographic symmetry (NCS) restraints for the ribosome as previously described[27]. The difference maps revealed the positions of the L1 stalk and domain 3 of the PSIV IRES, allowing us to position the models for these parts of the structure. The density corresponding to domain 1+2 of the IRES revealed the approximate positioning for this domain but was not sufficient to allow unambiguous building of the structural model. NCS-restrained structure refinement was carried out using PHENIX[32], as described[27]. Coot[33] was used for structure visualization and calculation of NCS-averaged maps. Figures were rendered using PyMOL[34]. Information on data collection and refinement statistics is summarized in Extended Data Table 1b.

31. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26,** 795–800 (1993).
32. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
33. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
34. DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, 2002).
35. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336,** 1030–1033 (2012).

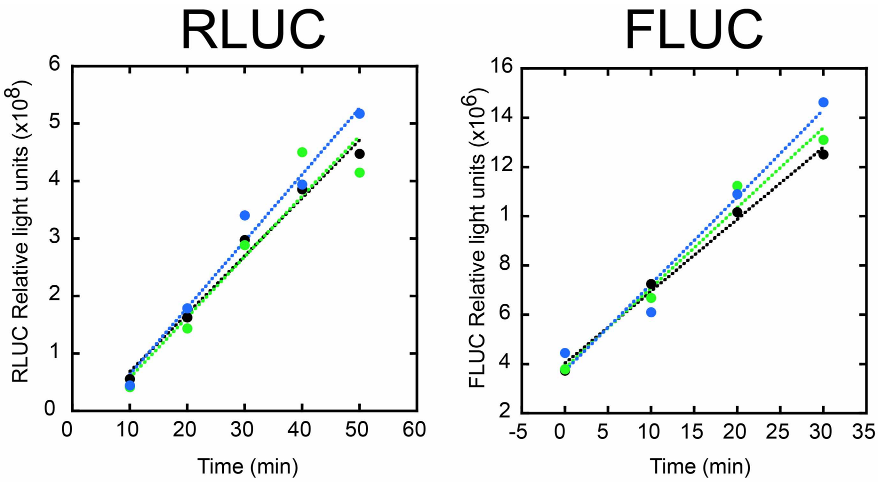


Extended Data Figure 1 | Canonical translation initiation signals, characteristics of IGR IRESs, and experimental design. **a**, Bacterial mRNAs (left) use a Shine–Dalgarno sequence (SDS; red) upstream of the AUG start codon and open reading frame (green) to recruit the 30S subunit (grey). The interaction is through the anti-SDS (A-SDS, yellow). Three initiation factors (magenta) are also important. Eukaryotic mRNAs (right) have a 5′ 7-methyl-guanosine 'cap' (red; 7 mG) that is bound by initiation factor 4E (4E, yellow). Multiple initiation factors (blue and magenta) serve to recruit the 40S subunit (grey) and allow it to scan to the start codon. **b**, Left, cryo-electron microscopy (cryo-EM) reconstruction of an IGR IRES RNA (magenta) bound to a human 40S subunit (yellow)[8]. The compact structure occupies the tRNA-binding groove of the subunit. Right, cryo-EM reconstruction of an IGR IRES RNA (magenta) bound to a human 80S ribosome[8]. The 40S subunit is yellow and the 60S subunit is cyan. The IRES RNA occupies the conserved intersubunit space. **c**, Cartoon representation of the secondary structure of a type 1 IGR IRES RNA (the type to which PSIV belongs). This structure is found between two open reading frames within the viral RNA genome. The two independently folded domains (domain 1+2 and domain 3) are indicated with dashed grey ovals. The locations of two pseudoknot interactions critical for inducing the correct IRES folded structure, and thus for function (PK 1 and PK 2), are shown. **d**, The structured IRES studied here is found in the intergenic region of the viral genome (red). It was placed into a dual-luciferase (LUC) reporter construct (blue, RLUC; yellow, FLUC) and this was cloned into bacterial expression vector pET30a. This vector was used to transform *Escherichia coli*. Induction of the culture leads to expression of the dual-LUC mRNA. Aliquots of the culture were harvested at defined time points and the amount of each LUC was measured. These data were used to determine the initial rate of LUC production (generally linear over the first 30–40 min post-induction) for each of the two reporters. RLUC served as a consistent internal control for different bacterial cultures, clones, growth rates, and so on.



Extended Data Figure 2 | Verification of independent quantifiable LUC production in bacteria. **a**, An empty pET30a vector (no inserted LUC reporter coding sequences) shows negligible signal. **b**, Traces of LUC activity as a function of time are shown from a construct in which the RLUC reporter was driven by the SDS and enhancer sequence from pET30a and FLUC was driven by an SDS only (Downstream SDS). The red octagon denotes stop codons. Both LUCs are generated, and RLUC production is higher, as expected. **c**, Removal of the SDS driving FLUC production (Downstream SDS_K/O) results in a loss of FLUC production, as expected. **d**, Insertion of the IRES from classical swine fever virus (CSFV) in a position to drive initiation of FLUC results in negligible FLUC activity. **a**–**d**, The *y*-axis indicates relative light units (RLU). Error bars represent 1 s.d. from the mean from three biological replicates. Here and throughout this study, we observed different LUC versus time profiles with different constructs. For example, the RLUC traces for the Downstream SDS and Downstream SDS_K/O constructs are different, despite no change to the SDS driving RLUC production (one shows a decrease of RLUC in later time points, the other maintains RLUC levels). The reason for this effect is unknown, but it only appears ~60 min after induction. **e**, Despite differences in longer time courses, LUC production was consistent and linear over the first 30–40 min post-induction. The RLUC and FLUC traces from the Downstream SDS and Downstream SDS_K/O constructs are shown. The consistency of these initial rates, before high levels of mRNA and reporter might build up and affect bacterial behaviour, justified their use as a means to quantitate LUC production (Extended Data Fig. 3).

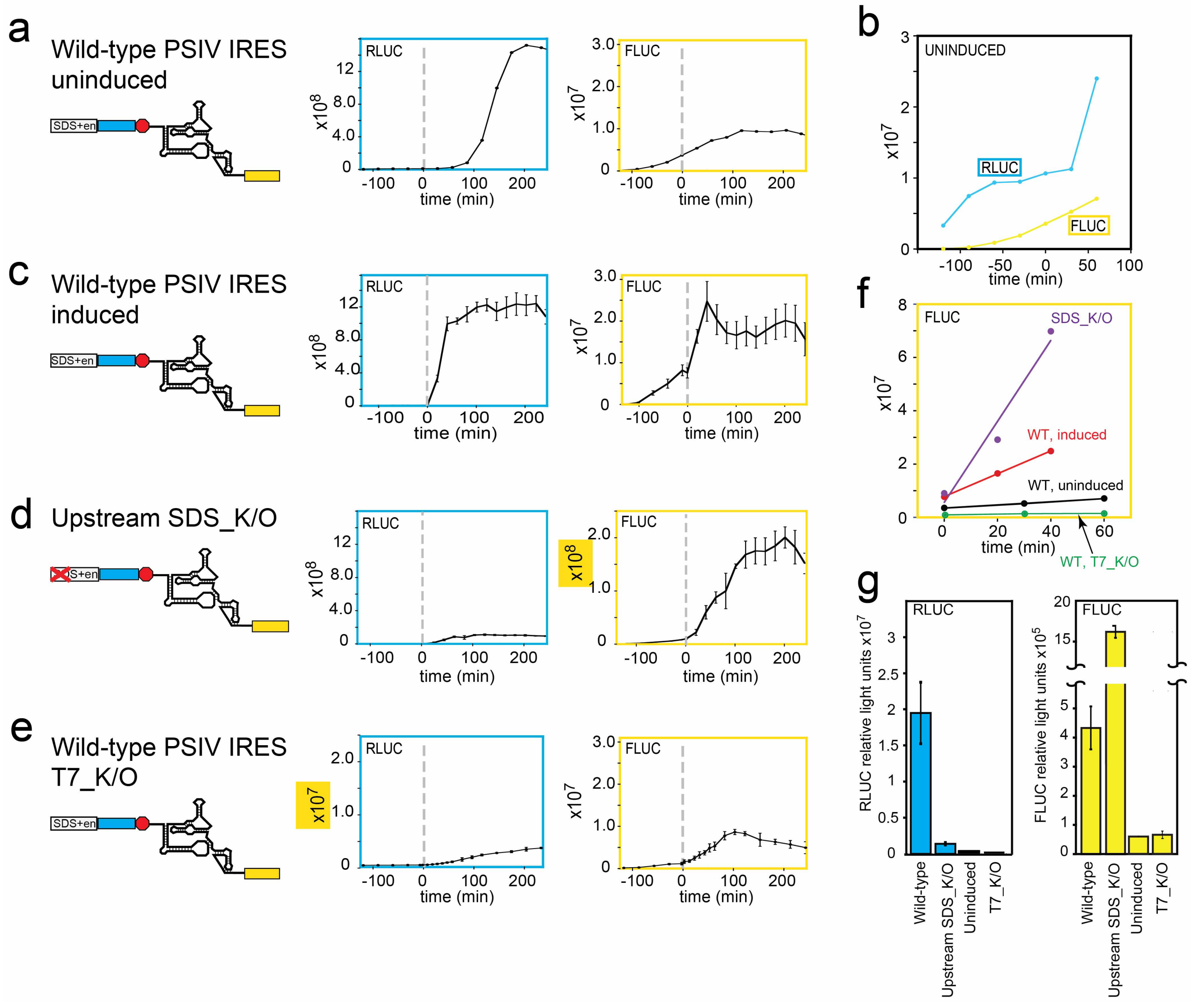
a



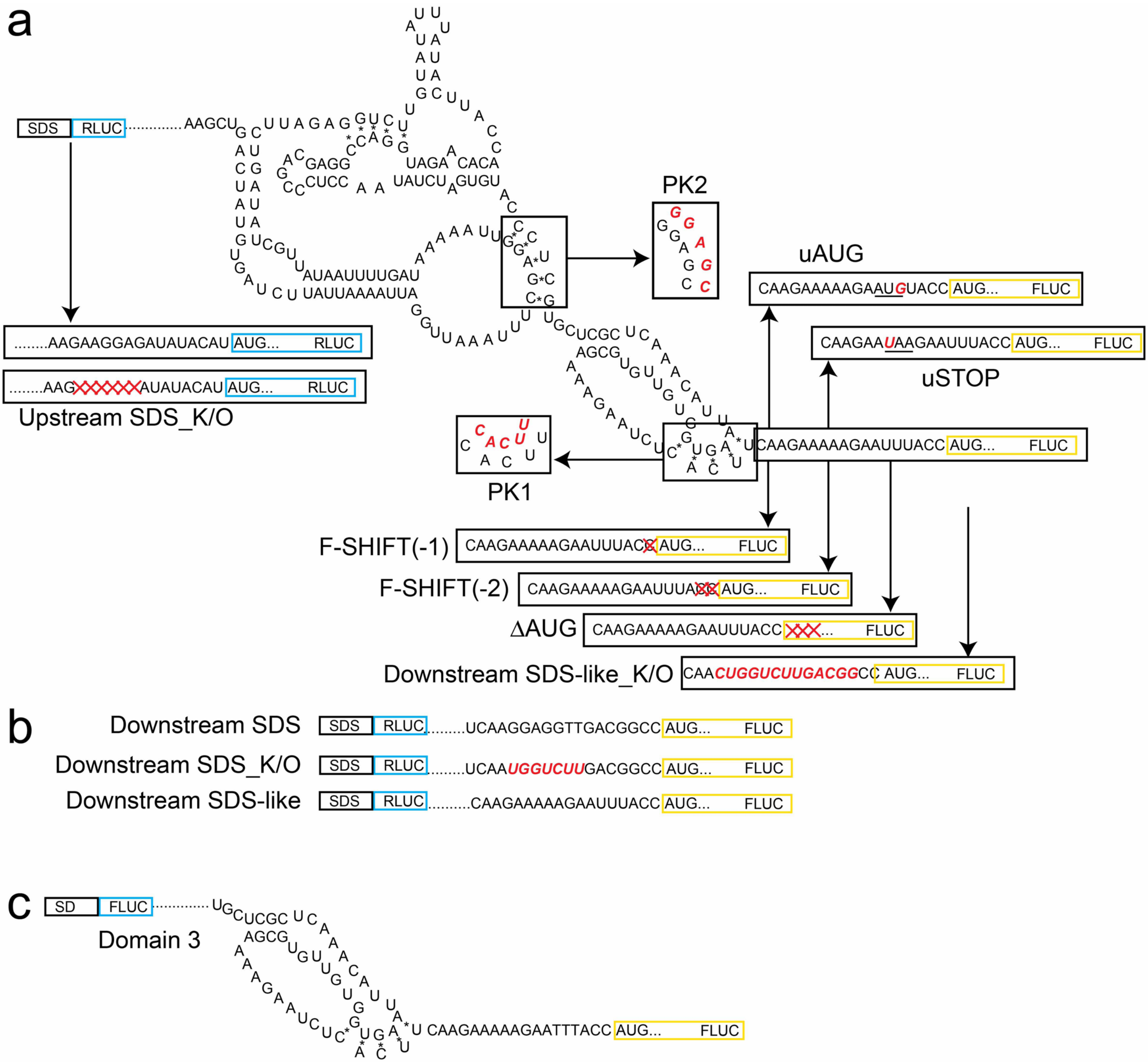
b

Run	RLUC initial rate	FLUC initial rate	Ratio (FLUC/RLUC)
1	9.14E+06	1.87E+05	2.05E-02
2	7.20E+06	1.30E+05	1.81E-02
3	8.49E+06	2.80E+05	3.30E-02
4	1.13E+07	2.92E+05	2.60E-02
5	1.37E+07	3.25E+05	2.37E-02
6	1.21E+07	3.53E+05	2.92E-02
AVERAGE:	1.03E+07	2.61E+05	2.51E-02
s.d.:	2.45E+06	8.52E+04	5.51E-03

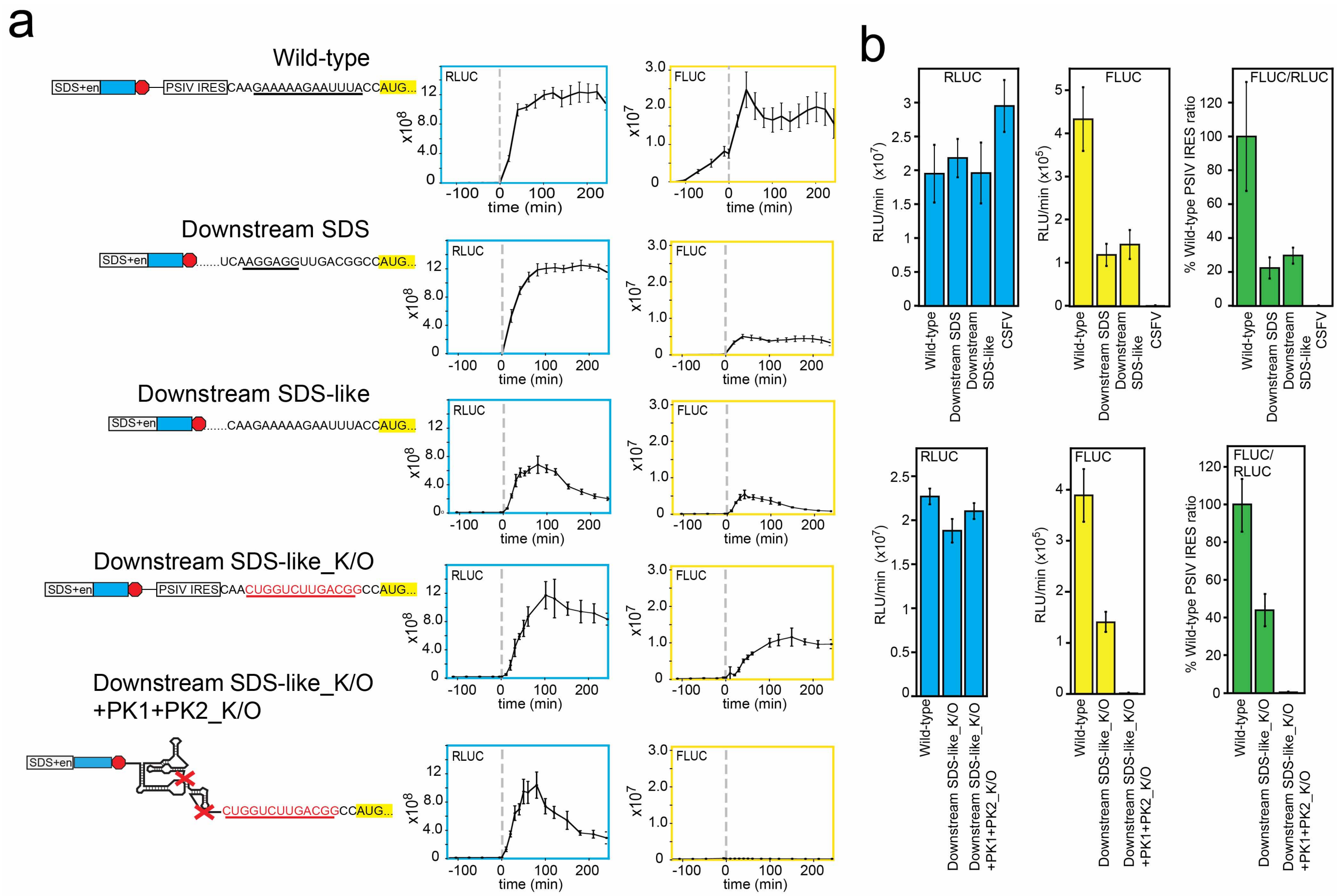
Extended Data Figure 3 | Determination of IRES activity from initial rates of LUC production. **a**, Representative graphs of RLUC and FLUC levels at early time points from three cultures of bacteria transformed with an IRES-containing bicistronic vector, induced with isopropyl-β-D-thiogalactoside (IPTG) at time = 0. Data from the three cultures are shown as black, green, and blue points, and a linear fit is shown with a dashed line for each. The slopes of these fit lines were used as the initial rate of LUC production per minute. **b**, Representative table of data for one IRES construct. Data from six cultures are shown, with initial rates for RLUC and FLUC production in RLU min^{-1}. Throughout this manuscript, the average rate for each LUC is shown in blue (RLUC) and yellow (FLUC) bar graphs. The ratio of these rates was determined from each culture, and these were averaged and shown in green bar graphs throughout the manuscript.



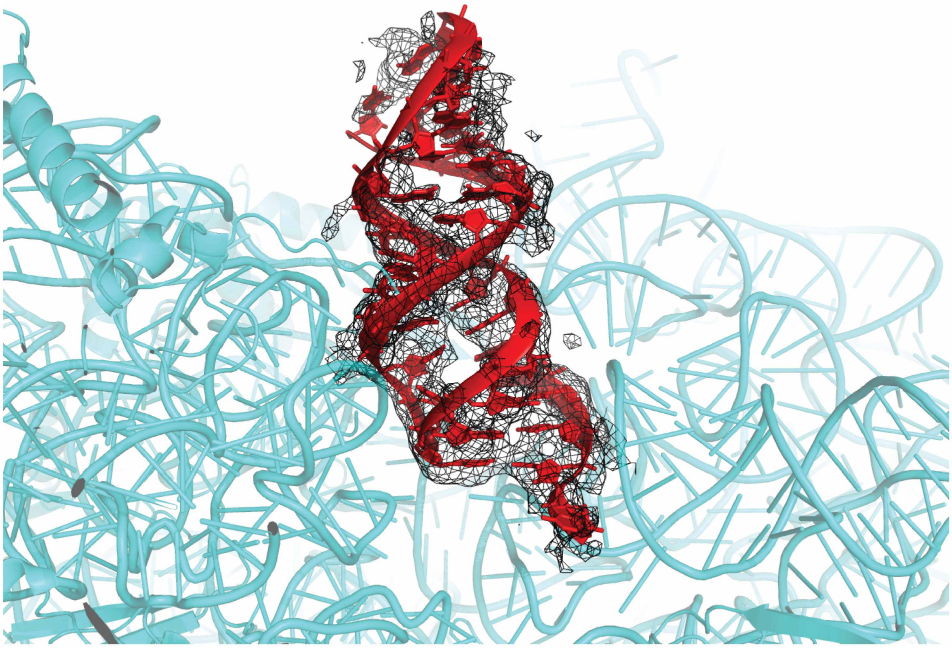
Extended Data Figure 4 | Examination of leaky expression and cryptic promoter activity. **a**, Traces of LUC production from the wild-type PSIV IRES-containing construct without induction with IPTG. Both RLUC and FLUC are produced due to 'leaky expression' of mRNA, a common observation with pET30a bacterial expression vectors. **a**–**f**, The *y*-axis shows RLU. **b**, Examination of the early time points of the traces from panel 1 show that both RLUC and FLUC are expressed to a low level without induction, and thus this leaky expression is not due to the IRES. **c**, Traces of wild-type PSIV IRES with IPTG induction at time = 0 (grey dashed line), showing the increase due to induction. **d**, Traces of a construct with the RLUC-driving SDS knocked out (Upstream SDS_K/O, same as in Fig. 1b), shown for comparison. **e**, To check for cryptic promoter activity due to transcription from a site other than the authentic T7 promoter, we cloned the full IRES-containing dual-LUC cassette into a pET30a vector in which the T7 promoter was mutated from 5′-TAATACGACTCACTATA-3′ to 5′-TAATGGTGTCTGAATTC-3′ (T7_K/O). Both RLUC and FLUC are produced to low levels, indicating some T7 promoter-independent expression exists in this vector, but the initial rates of producing upon induction are very low (see **f** and **g**). **f**, Initial rates of production of FLUC from the T7_K/O (induced), wild-type (uninduced), wild-type (induced), and Upstream SDS_K/O (induced) constructs. Rates of FLUC production from the T7_K/O and uninduced wild type are very low and not sufficient to account for apparent initiation from the IRES upon induction. This graph also illustrates the importance and utility of using the initial rates of LUC production for analysis, rather than the entire curve or an arbitrary later time point. **g**, Quantitated and graphed initial rate data for the four constructs in this figure. Error bars represent 1 s.d. of the mean from three biological replicates, except the uninduced control, which was done once.



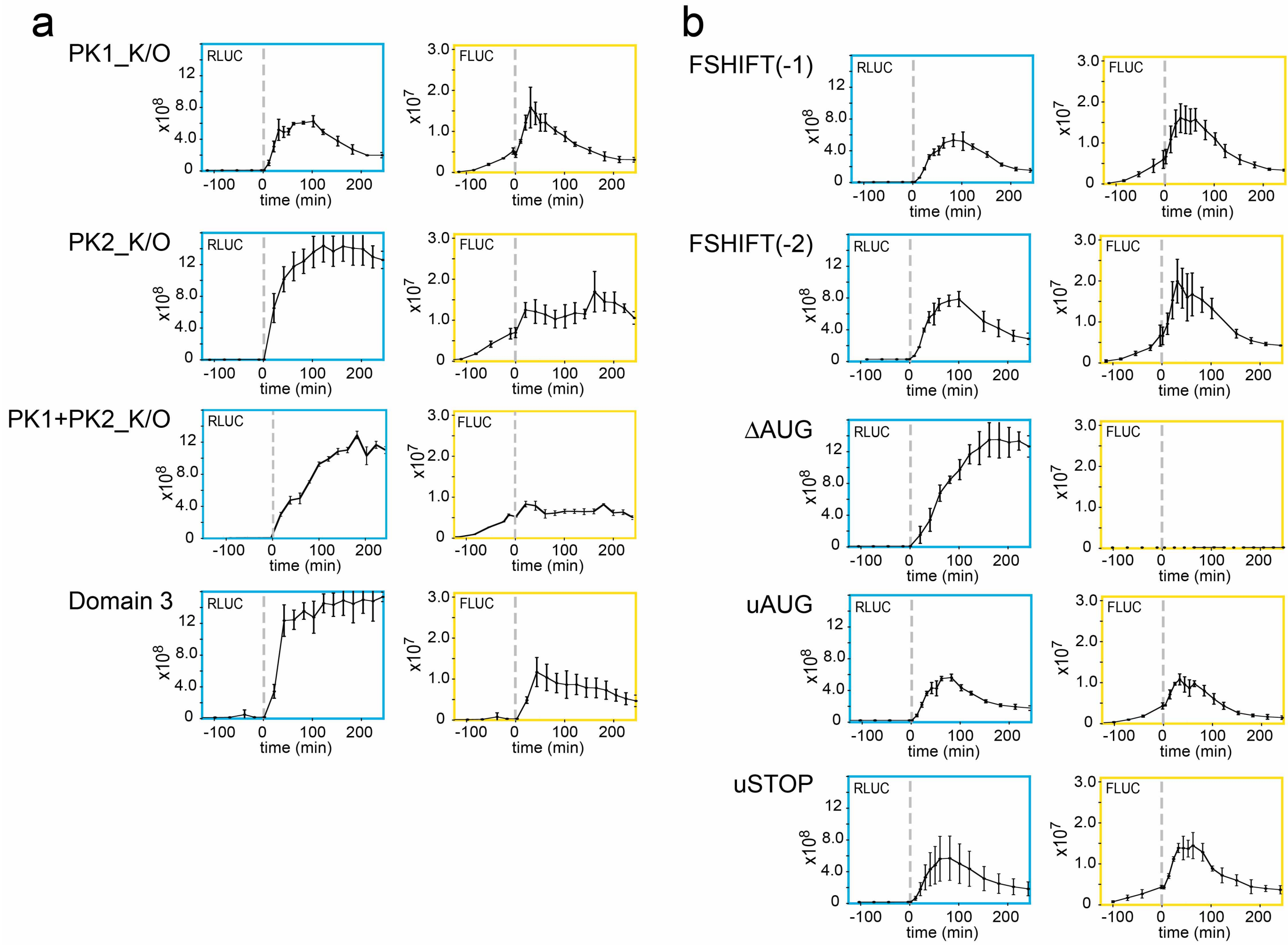
Extended Data Figure 5 | PSIV IGR IRES sequence, secondary structure, and design of mutants. **a**, Secondary structure of the full-length IGR IRES from the PSIV. The specific changes that were introduced to generate the mutants and constructs described and tested in the main text are shown. For each, the altered region is boxed and the change is shown in red. For the uAUG and uSTOP constructs, the start and stop codons are underlined. RLUC and FLUC coding sequences are boxed cyan and yellow, respectively. **b**, Constructs without the IRES that contain various wild-type or mutant SDS and SDS-like sequences upstream of the FLUC open reading frame. **c**, Construct containing just domain 3 of the PSIV IGR IRES.



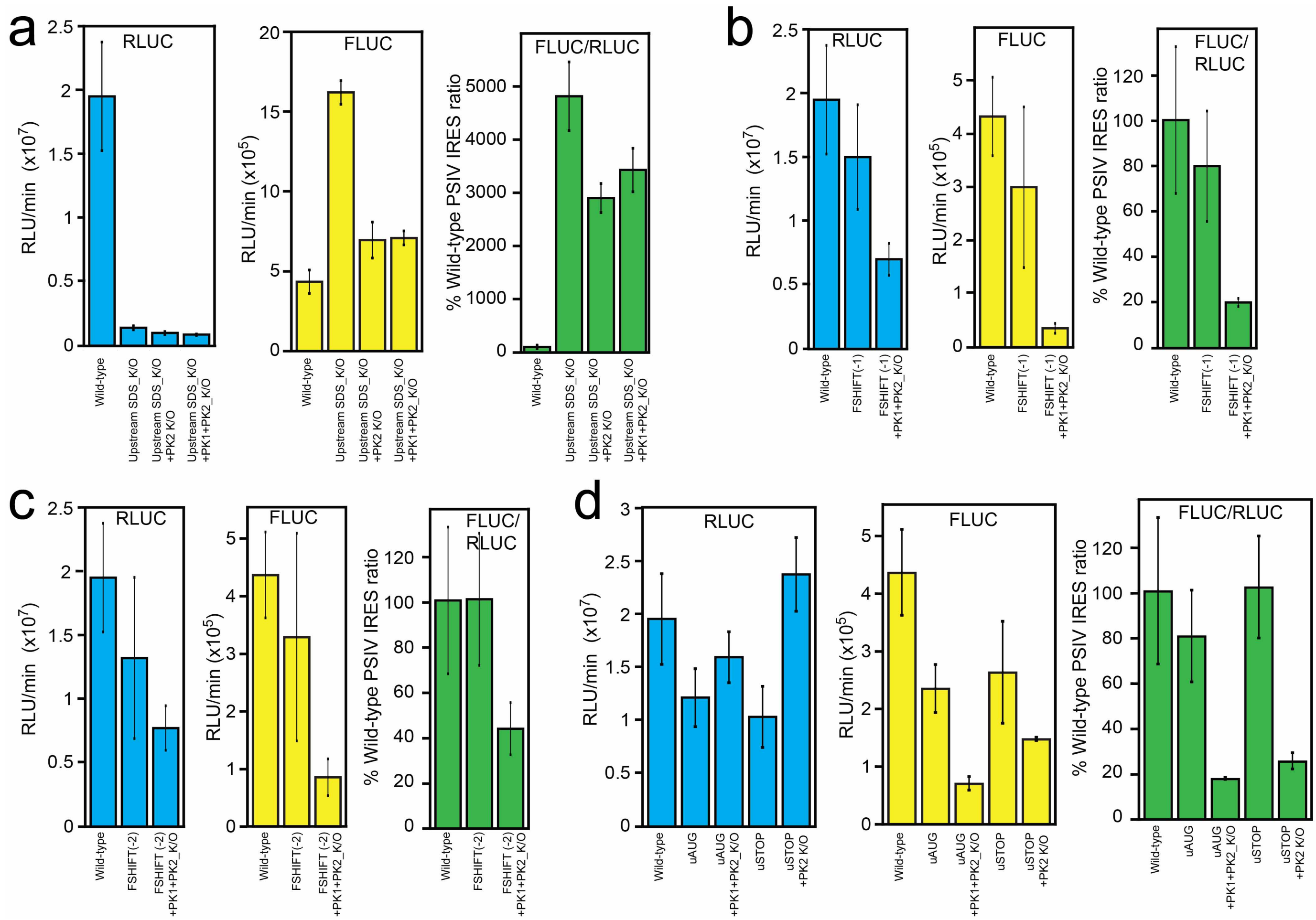
Extended Data Figure 6 | Contributions of region upstream of AUG to initiation activity. **a**, Diagram of constructs tested and traces of FLUC and RLUC production. The *y*-axis shows RLU. **b**, Quantitated initial rates from these constructs. Results from CSFV IRES (negative control) shown for comparison. 'Downstream SDS' contains an SDS driving FLUC production (in place of the IRES), 'Downstream SDS-like' contains the purine-rich sequence in place of the IRES and driving FLUC production. In 'Downstream SDS-like_K/O', this purine-rich sequence has been replaced by a pyrimidine-rich sequence. A PSIV IRES construct in which both pseudoknots are disrupted and the purine-rich SDS-like sequence just downstream of the IRES is mutated has essentially the same activity as the CSFV IRES (Downstream SDS-like_K/O+PK1+PK2_K/O). Error bars are 1 s.d. from the mean of three biological replicates.



Extended Data Figure 7 | The position of domain 3 in the full-length PSIV IGR IRES•70S structure. Crystal structure of a full-length PSIV IGR IRES bound to *T. thermophilus* 70S ribosomes. Cyan, small subunit; red, PSIV IRES domain 3; black, unbiased Fourier difference $F_o - F_c$ map for domain 3 in the P site of the small subunit. The large subunit and domain 1+2 are not shown.



Extended Data Figure 8 | Luciferase activity time courses for various constructs. a, Time-course traces for constructs and bar graphs shown in Fig. 3. **b**, Time-course traces for constructs and bar graphs shown in Fig. 4. Error bars are 1 s.d. from the mean of three biological replicates. **a**, **b**, The *y*-axis shows RLU.



Extended Data Figure 9 | Quantitated data for various constructs in the context of the PK1+PK2_K/O mutation. **a**, Combination of knocking out the RLUC SDS (Upstream SDS_K/O) with the PK2_K/O or PK1+PK2_K/O. Initial rates of RLUC are greatly diminished. Rates of FLUC are lower, but less diminished than RLUC. This is probably attributable to the decreased competition for ribosomes and the presence of the SDS-like sequence upstream of the FLUC open reading frame and not to robust initiation on the IRES. **b**, The PK1+PK2_K/O dramatically reduced the initial rate of FLUC production on the IRES with the F-SHIFT(−1) mutation. **c**, The PK1+PK2_K/O dramatically reduced the initial rate of FLUC production on the IRES with the F-SHIFT(−2) mutation. **d**, The PK1+PK2_K/O dramatically reduced the initial rate of FLUC production on the IRES with the uSTOP and uAUG mutations. Error bars are 1 s.d. from the mean from three biological replicates.

Extended Data Table 1 | Initial rates of RLUC and FLUC for all constructs tested and crystallographic data collection, phasing and refinement statistics

a

Construct or Condition Tested	FLUC (RLU/min; x 10^5)	RLUC (RLU/min; x 10^7)
T, uninduced	0.59	0.022
WT	4.33 ± 0.73	1.95 ± 0.42
WT, T7_K/O	0.65 ± 0.12	0.001 ± 0.0002
Upstream SDS_K/O	16.2 ± 0.73	0.14 ± 0.016
CSFV	0.0011 ± .0001	2.95 ± 0.38
Downstream SDS	1.18 ± 0.26	2.18 ± 0.28
PK1_K/O	2.84 ± 0.13	2.10 ± 0.63
PK2_K/O	2.67 ± 0.33	2.57 ± 0.40
Domain 3	2.82 ± 0.87	3.05 ± 0.21
PK1+PK2_K/O	0.66 ± 0.51	1.16 ± 0.21
ΔAUG	0.002 ± 0.0003	1.12 ± 0.17
uAUG	2.33 ± 0.41	1.21 ± 0.27
uSTOP	2.61 ± 0.88	1.03 ± 0.29
FSHIFT(-1)	3.00 ± 1.51	1.50 ± 0.41
FSHIFT(-2)	3.26 ± 1.79	1.32 ± 0.63
Downstream SDS-like	1.42 ± 0.34	1.96 ± 0.45
Downstream SDS-like_K/O	1.40 ± 0.19	1.88 ± 0.14
Downstream SDS-like_K/O +PK1+PK2_K/O	0.0013 ± 0.0012	2.10 ± 0.17
Upstream SDS_K/O +PK2_K/O	6.95 ± 1.13	0.097 ± 0.01
Upstream SDS_K/O +PK1+PK2_K/O	7.08 ± 0.44	0.084 ± 0.005
FSHIFT(-1)+PK1+PK2_K/O	0.35 ± 0.092	0.70 ± 0.12
FSHIFT(-2)+PK1+PK2_K/O	0.84 ± 0.32	0.72 ± 0.17
uAUG+PK1+PK2_K/O	0.69 ± 0.11	1.59 ± 0.24
uSTOP+PK1+PK2_K/O	1.46 ± 0.03	2.37 ± 0.35

b

Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	209.05, 447.22, 608.96
α, β, γ (°)	90, 90, 90
Resolution (Å)	60-3.8 (3.8 – 4.0)*
R_{meas}†	0.2 (1.6)
CC(1/2)‡	99.5 (41.6)
$I / \sigma I$	8.28 (1.2)
Completeness (%)	99.9 (99.8)
Redundancy	4.8 (3.4)
Refinement	
Resolution (Å)	60 – 3.8
No. reflections	555,726
R_{work} / R_{free}	0.246/0.284
No. atoms	287428
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.702

a, Raw values are shown for all constructs tested. All values are the mean of three independent experiments ± 1 s.d. from the mean, except for the uninduced control that was done once. WT, wild type.
b, Crystallographic statistics.
* Values in parentheses are for highest-resolution shell.
† R_{meas} is R_{meas} as reported by XDS[31].
‡ CC(1/2) is the percentage of correlation between intensities from random half-data sets as defined previously[35].

doi:10.1038/nature14033

Structure of the F-actin–tropomyosin complex

Julian von der Ecken[1], Mirco Müller[2], William Lehman[3], Dietmar J. Manstein[2], Pawel A. Penczek[4] & Stefan Raunser[1]

Filamentous actin (F-actin) is the major protein of muscle thin filaments, and actin microfilaments are the main component of the eukaryotic cytoskeleton. Mutations in different actin isoforms lead to early-onset autosomal dominant non-syndromic hearing loss[1], familial thoracic aortic aneurysms and dissections[2], and multiple variations of myopathies[3]. In striated muscle fibres, the binding of myosin motors to actin filaments is mainly regulated by tropomyosin and troponin[4,5]. Tropomyosin also binds to F-actin in smooth muscle and in non-muscle cells and stabilizes and regulates the filaments there in the absence of troponin[6]. Although crystal structures for monomeric actin (G-actin) are available[7], a high-resolution structure of F-actin is still missing, hampering our understanding of how disease-causing mutations affect the function of thin muscle filaments and microfilaments. Here we report the three-dimensional structure of F-actin at a resolution of 3.7 Å in complex with tropomyosin at a resolution of 6.5 Å, determined by electron cryomicroscopy. The structure reveals that the D-loop is ordered and acts as a central region for hydrophobic and electrostatic interactions that stabilize the F-actin filament. We clearly identify map density corresponding to ADP and Mg^{2+} and explain the possible effect of prominent disease-causing mutants. A comparison of F-actin with G-actin reveals the conformational changes during filament formation and identifies the D-loop as their key mediator. We also confirm that negatively charged tropomyosin interacts with a positively charged groove on F-actin. Comparison of the position of tropomyosin in F-actin–tropomyosin with its position in our previously determined F-actin–tropomyosin–myosin structure[8] reveals a myosin-induced transition of tropomyosin. Our results allow us to understand the role of individual mutations in the genesis of actin- and tropomyosin-related diseases and will serve as a strong foundation for the targeted development of drugs.

To determine the structure of F-actin is inherently difficult because of its flexibility and its resistance to crystallization. Therefore, the only structural models of F-actin so far have been determined either from medium-resolution electron cryomicroscopy (cryo-EM) maps[9–13] or by interpreting X-ray fibre diffraction data[14], which has certain limitations. Using a direct electron detector and drift correction and by improving the image processing of helical specimens (see Methods), we have determined the structure of F-actin in complex with tropomyosin at an average resolution of 3.7 Å for F-actin and 6.5 Å for tropomyosin using cryo-EM (Fig. 1a, Extended Data Figs 1a, b, 2a, Supplementary Video 1). During refinement the helical parameters—that is, the rise per subunit and the azimuthal rotation—were estimated to be 27.5 Å and 166.4°, respectively (see Methods). The side-chain densities of most actin residues were clearly resolved (Extended Data Fig. 3, Supplementary Video 2) and allowed us to build an atomic model of F-actin (Fig. 1b, Extended Data Fig. 3). The first four residues of the amino terminus and the last four residues of the carboxy terminus were not resolved (Extended Data Fig. 2b–d), indicating that these regions are disordered in the filament. However, we could clearly identify map density corresponding to ADP and the coordinated cation, which is most probably Mg^{2+} (Fig. 1b, Extended Data Fig. 3a).

The overall organization of F-actin is similar to that described in previous structures and models[10,14]. However, given the superior resolution of our structure, we could clearly identify many salt bridges and therefore directly reveal intra- and intermolecular interactions of the F-actin filament in detail (Extended Data Fig. 4a).

F-actin is composed of two long-pitch helical strands. Interactions between actin subunits of the same strand and the opposing strand—the so-called intrastrand and interstrand interactions, respectively—stabilize the F-actin filament (Fig. 2, Extended Data Figs 4 and 5). Intrastrand contacts are mediated by subdomains SD2 and SD4 of one actin subunit with the SD3 of the adjacent actin subunit of the same strand (Fig. 2a, b). Besides several salt bridges between the edges of SD4 and SD3 (Fig. 2b), the major site of interstrand interaction is between the D-loop and the bottom of the β-sheet of SD3 (Fig. 2c–f, Extended Data Fig. 4b–e). The D-loop encloses tyrosine 169 of the neighbouring subunit, resembling a lock-and-key interaction (Fig. 2c). In addition, adjacent residues fit snugly into the groove formed by regions next to the D-loop around isoleucine 64 (Fig. 2d) and a prominent hydrophobic patch in the D-loop interacts with a hydrophobic groove on the neighbouring F-actin subunit (Fig. 2e, f). Thus, the intrastrand interface at this position is large and is stabilized not only by electrostatic and hydrophobic interactions but also by geometric surface complementarity. As expected, mutations at this site (which, for instance, induce charges at hydrophobic interfaces) hamper F-actin polymerization and stability (Extended Data Fig. 4d, e) and can lead to different forms of nemaline myopathy[3,15] (Extended Data Fig. 4a).

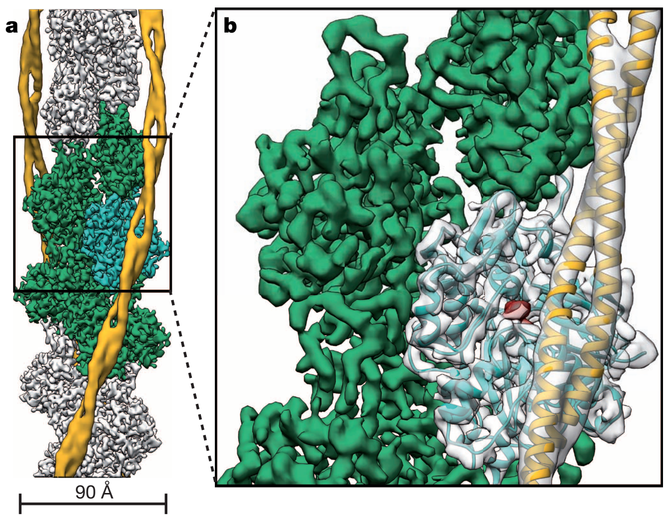


Figure 1 | Cryo-EM structure of F-actin decorated with tropomyosin. a, Full cryo-EM reconstruction of F-actin (grey, with five central subunits in green and one subunit in cyan) decorated with tropomyosin (yellow). **b**, Close-up view of **a** with the atomic and molecular model of an F-actin subunit (cyan) and tropomyosin (yellow) and their corresponding densities, respectively. The density corresponding to ADP is depicted in red.

[1]Department of Structural Biochemistry, Max Planck Institute of Molecular Physiology, 44227 Dortmund, Germany. [2]Institute for Biophysical Chemistry, Hannover Medical School, 30625 Hannover, Germany. [3]Department of Physiology and Biophysics, Boston University School of Medicine, Boston, Massachusetts 02118, USA. [4]Department of Biochemistry and Molecular Biology, The University of Texas, Houston Medical School, Houston, Texas 77030, USA.

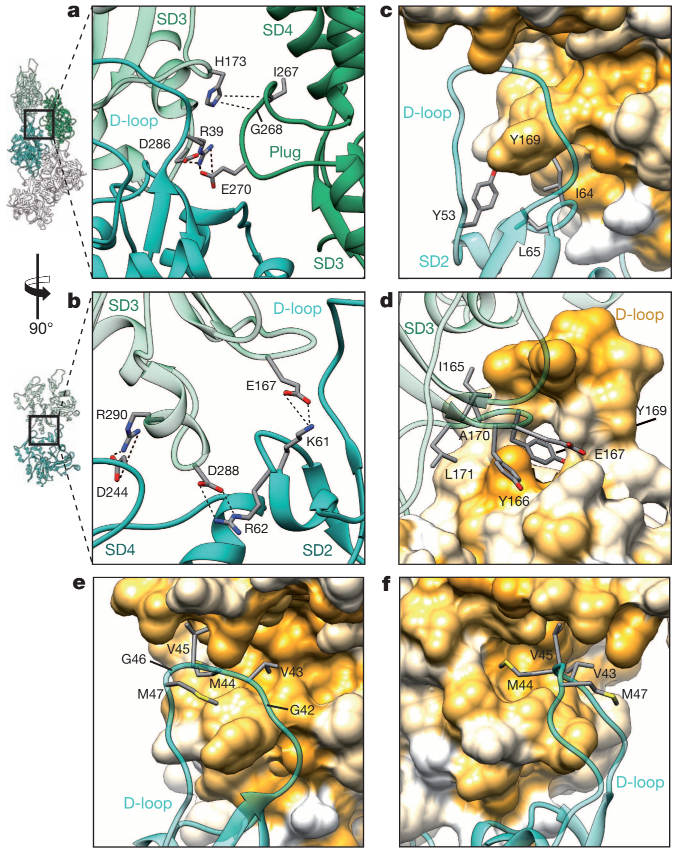


Figure 2 | Filament stability by intrastrand and interstrand interactions. **a**–**f**, Neighbouring F-actin subunits stabilize the F-actin filament by interaction through salt bridges (**a**, **b**) and by hydrophobic interactions (**c**–**f**). The central F-actin subunit is depicted in cyan, and adjacent subunits are shown in shades of green. Surfaces are coloured from high (yellow) to low (white) hydrophobicity. Interactions between amino acids are highlighted with dotted lines. **a**, Interstrand and intrastrand salt bridges of three actin subunits involving the plug. **b**, Several intrastrand salt bridges at the actin–actin interface. **c**–**f**, Front view (**c**, **e**), back view (**d**) and side view (**f**) of the D-loop interacting with the SD3 of the neighbouring intrastrand F-actin subunit. The D-loop wraps around Y169 of the neighbouring subunit and other residues snugly fit into the groove formed by regions adjacent to the D-loop (**c**, **d**), resembling a lock-and-key interaction. In addition, a prominent hydrophobic patch in the D-loop interacts with a hydrophobic groove on the neighbouring F-actin subunit (**e**, **f**).

In contrast to most G-actin structural models, in which the D-loop is not resolved owing to its high flexibility, the D-loop is well defined in our structure (Extended Data Fig. 3b), corroborating its importance in mediating actin–actin interaction. It also contrasts with results of medium-resolution cryo-EM data[12] and disulphide cross-linking studies[16] that suggested the presence of multiple D-loop conformational states. In addition, these and other studies[17,18] suggested a close interaction of the D-loop with the C terminus of actin. Although the C terminus is located next to the D-loop, it is not resolved in the cryo-EM density (Extended Data Fig. 2c), indicating that the interaction between these two regions is either transient or structurally flexible.

Our structural determination also gives a possible explanation for the recently described F-actin disassembly by Mical, a multidomain cytosolic actin regulator[19]. Mical post-translationally oxidizes methionine 44 within the D-loop of actin[20], leading to the less flexible and more hydrophilic sulfoxide. The increased polarity could lead to a weakening of the hydrophobic actin–actin interaction (Extended Data Fig. 4d) and consequently to a Mical-induced disassembly of the F-actin filament.

The F-actin structure confirms that the so-called hydrophobic plug[9] connects three F-actin subunits with each other (Fig. 2a). However, as suggested previously[10], this interstrand interaction is formed by salt bridges (Fig. 2a), not by hydrophobic interactions (Extended Data Fig. 5a–e). A large part of the plug is negatively charged on the surface, electrostatically interacting with a positively charged patch on the opposing actin, with histidine 173 and lysine 284 at its centre (Extended Data Fig. 5c–e). Another possible salt bridge connects two neighbouring actins at lysine 113 of SD1 and glutamate 195 of SD4 (Extended Data Fig. 5f). However, the interface at this position is quite small and not surrounded by other prominent electrostatic or hydrophobic interactions. In general, the two strands of F-actin are held together by only a few contacts, and a large empty space is maintained between the strands. This is probably an important requisite for the flexibility of F-actin filaments.

As previously described[9,10,14], F-actin is more flattened than is G-actin because of a rotation of SD1 and particularly SD2 (Fig. 3a, Supplementary Video 3). New salt bridges between SD1, SD4 and SD2 stabilize the new interface between the opposing subdomains at the nucleotide-binding cleft (Fig. 3b–f, Extended Data Fig. 6a).

The coordination of ADP and the divalent cation in the nucleotide-binding cleft by nearby residues is very similar to that in G-actin (Fig. 3b, Extended Data Fig. 6b). However, glutamine 137, which has been shown to play a central role in ATP hydrolysis[21], is clearly moved closer to ADP, coordinating not only the cation (as in G-actin) but also the nucleotide

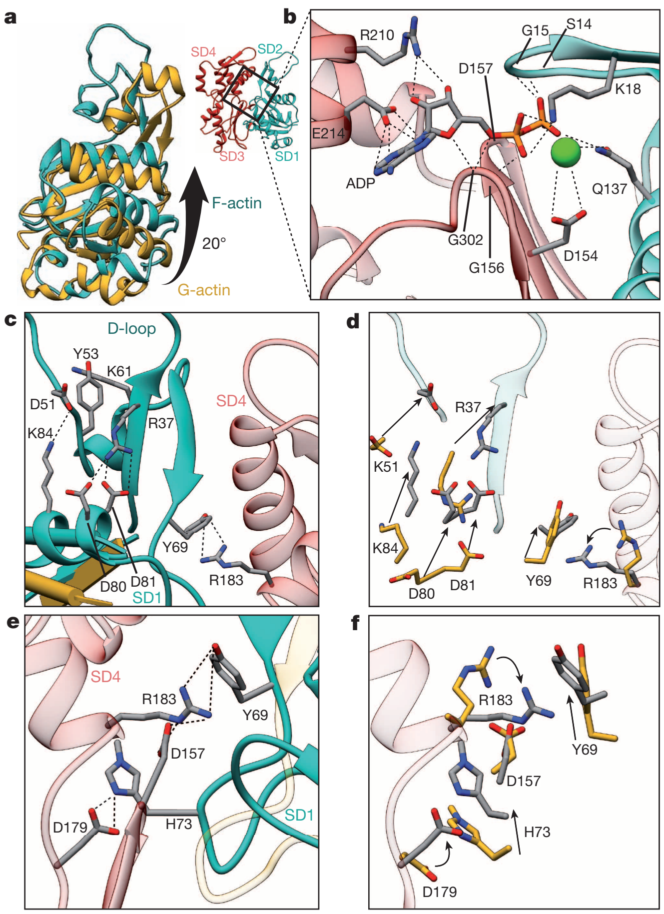


Figure 3 | G-actin to F-actin transition. **a**, A global rotation of SD1 and SD2 leads to a flattening of G-actin (yellow) during transition to F-actin (cyan). **b**, Coordination of ADP and a cation in the nucleotide binding cleft in F-actin (SD1–SD2 cyan, SD3–SD4 red). The cation is most probably Mg^{2+}. However, Ca^{2+} cannot be excluded as the coordinated cation because the actin was purified in Ca^{2+}-containing buffer. **c**–**f**, The conformation of F-actin is finally stabilized by intramolecular interactions between SD2, SD1 and SD4 (**c**, **e**). Most of the residues involved in the stabilization of the F-actin conformation show a considerable movement during the transition between G-actin and F-actin (**d**, **f**). **a**, G-actin with defined D-loop for better visualization; PDB accession code 1J6Z. **c**–**f**, ATP-bound G-actin; PDB accession code 3EL2.

β-phosphate (Extended Data Fig. 6c, d, Supplementary Video 3). Interestingly, the presence of ATP instead of ADP in the nucleotide-binding site would be sterically unfavourable (Extended Data Fig. 6d, Supplementary Video 3), suggesting that a different intermediate conformation exists for F-actin–ATP. High-resolution crystal structures of non-vertebrate actin showed that glutamine 137 positions the nucleophilic water that attacks the γ-phosphate of ATP[22]. Therefore, as previously suggested[14], the shorter distance of glutamine 137 to the γ-phosphate probably induces ATP hydrolysis and then afterwards the cation takes the position of the γ-phosphate in the finally resulting ADP-binding-state (Extended Data Fig. 6e, Supplementary Video 3). A mutation of glutamine 137 connected to nemaline myopathy[23] further highlights the importance of this residue in the proper function of F-actin (Extended Data Fig. 6f).

The comparison of the structure of G-actin (PDB accession codes 3EL2 and 1J6Z) with that of F-actin, which represent the start and end points of filament formation, allows us to describe the possible conformational changes during polymerization (Extended Data Fig. 7, Supplementary Discussion).

The F-actin filaments were completely decorated by tropomyosin (Fig. 1a, Extended Data Fig. 1a, b, i). Although tropomyosin consists of seven pseudo-repeating units, applying the helical symmetry of F-actin for the reconstruction of the F-actin–tropomyosin complex results in a reduced resolution of tropomyosin and the overlap region of the N and C termini is not visible in the map (see Methods). However, densities corresponding to the helices of the coiled-coil were well resolved (Fig. 1a, b) and allowed us to fit a previous chimeric model of tropomyosin derived from X-ray, nuclear magnetic resonance and molecular dynamics simulation data[24] and to determine the register of the chains of tropomyosin. Together with the atomic resolution of F-actin, this allows us to describe the interaction between actin and tropomyosin in greater detail and with higher accuracy than previous studies.

Tropomyosin, which is mainly negatively charged on its surface, interacts with a positively charged groove on F-actin, confirming previous predictions (compare refs 24–26) (Fig. 4a, b, Extended Data Fig. 8a). In particular, lysines 326 and 328 on F-actin strongly interact with the different pseudo-repeats of tropomyosin (Fig. 4c). The distance between tropomyosin and actin varies depending on the position along the filament between 38 Å and 40 Å. The distance would allow for direct interactions

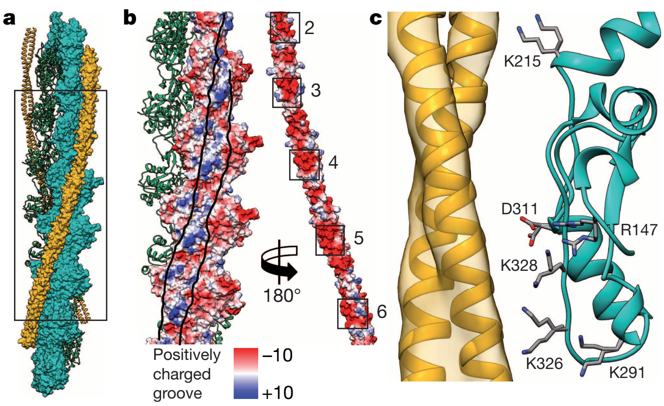


Figure 4 | F-actin interaction with tropomyosin. **a**, Structural overview of an F-actin filament (green and cyan) decorated with tropomyosin (yellow). Half of the filament is shown in surface representation. The N and C termini of F-actin are included in this model. **b**, Surface of F-actin and tropomyosin (pseudo-repeats 2–6) with the electrostatic Coulomb potentials ranging from $-10\,\mathrm{kcal\,mol^{-1}}$ to $+10\,\mathrm{kcal\,mol^{-1}}$ at pH 7.5. Tropomyosin was rotated by 180° and shifted to the right to allow a better view on the F-actin–tropomyosin interface, which is delimited by lines drawn onto the surfaces. The overall negatively charged tropomyosin interacts with a positively charged groove on F-actin. **c**, Several charged residues of actin are within distances that would make it possible to interact with tropomyosin via putative salt bridges. Different rotamers of the same residue are shown to indicate how F-actin subunits could adjust to the surfaces of different tropomyosin pseudo-repeats.

of opposing residues (Fig. 4c) and is similar to the distance in the M-state[8] and to what was previously suggested for the other states of tropomyosin[25,27]. The interactions between tropomyosin and F-actin probably determine the position of tropomyosin on F-actin in the Apo-state (A-state), a term we recently introduced to describe the position of tropomyosin in the absence of troponin, myosin or Ca^{2+} (ref. 26).

The position of tropomyosin in the A-state in our structure (Fig. 4a, Extended Data Fig. 8b) is close to that described by Sousa *et al.*[13], but differs from a previous positioning based on negative stain data and molecular dynamics simulations[24]. To determine the source of this discrepancy, we first repeated the negative stain work (Extended Data Fig. 1c–j) and indeed reproduced the different position (Extended Data Fig. 8c). However, when we cross-linked the sample before staining, tropomyosin was in the same position as in our cryo-EM structure (Extended Data Fig. 1c–h, j, Extended Data Fig. 8d–g). This shows that negative staining rather than cryo-preparation results in a change of the tropomyosin position on F-actin. The relatively low pH of the stain (pH 4) probably induces repulsions at the F-actin–tropomyosin interface and forces tropomyosin to change its position (Extended Data Fig. 8h). In any case, the change in position depending on preparative conditions confirms the striking ability of tropomyosin to translocate azimuthally across the actin filament at low energy cost[26].

In comparison to the previously assumed position of tropomyosin in the A-state, the new position blocks only minor regions of the myosin-binding site on F-actin. This supports our previous model of myosin binding to actin filaments in the absence of troponin[8] even more strongly. If the current position of tropomyosin is the default state for tropomyosin, then only loop 4 and the cardiomyopathy loop of myosin would be sterically hindered from binding to the F-actin filament (Extended Data Fig. 9a–d, Supplementary Video 4) and probably allows for a weak binding of myosin, especially with its lower 50-kDa domain. Actin-induced cleft closure of myosin would then move tropomyosin to its M-state[8] (Extended Data Fig. 9b–g, Supplementary Video 4).

Comparing the position of tropomyosin in the new A-state with the M-state (Extended Data Fig. 9h), we find two possible transitions between the states. Either tropomyosin rolls on the F-actin filament by an azimuthal rotation of about 16° and a left-handed rotation of about 70° or a right-handed rotation of about 110° along its own axis (Extended Data Fig. 9i, j) or tropomyosin shifts by about 37 Å (half the length of a pseudo-repeat) along the actin filament (Extended Data Fig. 9k, l) or a mixture of the two effects occurs. The distance of tropomyosin to the filament axis of F-actin is not altered during the transition. Gestalt-binding and rigidity of tropomyosin[28] support a shifting mechanism, whereas the spatial limitations inside the sarcomere make a shift of 37 Å rather unlikely. However, direct localization of the N/C-terminal overlap of tropomyosin will be needed to discriminate between the two possibilities.

In summary, the structure of the F-actin-tropomyosin complex shows how F-actin filaments are stabilized in health and destabilized in certain diseases. Our results provide the structural framework for further experiments towards resolving the molecular details of actin polymerization and its interaction with actin-regulating proteins such as tropomyosin.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 August; accepted 7 November 2014.
Published online 1 December 2014.

1. van Wijk, E. *et al.* A mutation in the gamma actin 1 (ACTG1) gene causes autosomal dominant hearing loss (DFNA20/26). *J. Med. Genet.* **40,** 879–884 (2003).
2. Milewicz, D. M. *et al.* Genetic basis of thoracic aortic aneurysms and dissections: focus on smooth muscle cell contractile dysfunction. *Annu. Rev. Genomics Hum. Genet.* **9,** 283–302 (2008).
3. Sparrow, J. C. *et al.* Muscle disease caused by mutations in the skeletal muscle alpha-actin gene (ACTA1). *Neuromuscul. Disord.* **13,** 519–531 (2003).
4. Gordon, A. M., Homsher, E. & Regnier, M. Regulation of contraction in striated muscle. *Physiol. Rev.* **80,** 853–924 (2000).

5. Maytum, R., Lehrer, S. S. & Geeves, M. A. Cooperativity and switching within the three-state model of muscle regulation. *Biochemistry* **38,** 1102–1110 (1999).
6. Pittenger, M. F., Kazzaz, J. A. & Helfman, D. M. Functional properties of non-muscle tropomyosin isoforms. *Curr. Opin. Cell Biol.* **6,** 96–104 (1994).
7. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. Atomic structure of the actin:DNase I complex. *Nature* **347,** 37–44 (1990).
8. Behrmann, E. *et al.* Structure of the rigor actin-tropomyosin-myosin complex. *Cell* **150,** 327–338 (2012).
9. Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. Atomic model of the actin filament. *Nature* **347,** 44–49 (1990).
10. Fujii, T., Iwane, A. H., Yanagida, T. & Namba, K. Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature* **467,** 724–728 (2010).
11. Murakami, K. *et al.* Structural basis for actin assembly, activation of ATP hydrolysis, and delayed phosphate release. *Cell* **143,** 275–287 (2010).
12. Galkin, V. E., Orlova, A., Schröder, G. F. & Egelman, E. H. Structural polymorphism in F-actin. *Nature Struct. Mol. Biol.* **17,** 1318–1323 (2010).
13. Sousa, D. R., Stagg, S. M. & Stroupe, M. E. Cryo-EM structures of the actin:tropomyosin filament reveal the mechanism for the transition from C- to M-state. *J. Mol. Biol.* **425,** 4544–4555 (2013).
14. Oda, T., Iwasa, M., Aihara, T., Maéda, Y. & Narita, A. The nature of the globular- to fibrous-actin transition. *Nature* **457,** 441–445 (2009).
15. Laing, N. G. *et al.* Mutations and polymorphisms of the skeletal muscle alpha-actin gene (ACTA1). *Hum. Mutat.* **30,** 1267–1277 (2009).
16. Öztug Durer, Z. A., Diraviyam, K., Sept, D., Kudryashov, D. S. & Reisler, E. F-actin structure destabilization and DNase I binding loop fluctuations: mutational cross-linking and electron microscopy analysis of loop states and effects on F-actin. *J. Mol. Biol.* **395,** 544–557 (2010).
17. Strzelecka-Gołaszewska, H., Mossakowska, M., Woźniak, A., Moraczewska, J. & Nakayama, H. Long-range conformational effects of proteolytic removal of the last three residues of actin. *Biochem. J.* **307,** 527–534 (1995).
18. Crosbie, R. H. *et al.* Structural connectivity in actin: effect of C-terminal modifications on the properties of actin. *Biophys. J.* **67,** 1957–1964 (1994).
19. Hung, R.-J. *et al.* Mical links semaphorins to F-actin disassembly. *Nature* **463,** 823–827 (2010).
20. Hung, R.-J., Pak, C. W. & Terman, J. R. Direct redox regulation of F-actin assembly and disassembly by Mical. *Science* **334,** 1710–1713 (2011).
21. Iwasa, M., Maeda, K., Narita, A., Maéda, Y. & Oda, T. Dual roles of Gln137 of actin revealed by recombinant human cardiac muscle alpha-actin mutants. *J. Biol. Chem.* **283,** 21045–21053 (2008).
22. Vorobiev, S. *et al.* The structure of nonvertebrate actin: implications for the ATP hydrolytic mechanism. *Proc. Natl Acad. Sci. USA* **100,** 5760–5765 (2003).
23. Koy, A. *et al.* Nemaline myopathy with exclusively intranuclear rods and a novel mutation in ACTA1 (Q139H). *Neuropediatrics* **38,** 282–286 (2007).
24. Li, X. E. *et al.* Tropomyosin position on F-actin revealed by EM reconstruction and computational chemistry. *Biophys. J.* **100,** 1005–1013 (2011).
25. Lorenz, M., Poole, K. J., Popp, D., Rosenbaum, G. & Holmes, K. C. An atomic model of the unregulated thin filament obtained by X-ray fiber diffraction on oriented actin-tropomyosin gels. *J. Mol. Biol.* **246,** 108–119 (1995).
26. Lehman, W., Orzechowski, M., Li, X. E., Fischer, S. & Raunser, S. Gestalt-binding of tropomyosin on actin during thin filament activation. *J. Muscle Res. Cell Motil.* **34,** 155–163 (2013).
27. Poole, K. J. V. *et al.* A comparison of muscle thin filament models obtained from electron microscopy reconstructions and low-angle X-ray fibre diagrams from non-overlap muscle. *J. Struct. Biol.* **155,** 273–284 (2006).
28. Holmes, K. C. & Lehman, W. Gestalt-binding of tropomyosin to actin filaments. *J. Muscle Res. Cell Motil.* **29,** 213–219 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank O. Hofnagel for excellent assistance in cryo sample preparation and electron microscopy, R. S. Goody for continuous support and for comments on the manuscript and I. Vetter for assistance in data processing. We gratefully acknowledge R. Matadeen and S. de Carlo (FEI Company) for image acquisition at the National Center for Electron Nanoscopy in Leiden (NeCEN), which is co-financed by grants from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (project 175.010.2009.001) and by the European Union's Regional Development Fund through 'Kansen voor West' (project 21Z.014). J.v.d.E. is a fellow of Studienstiftung des deutschen Volkes. This work was supported by the Behrens-Weise foundation (to S.R.), NIH U54 094598 and R01 60635 (to P.A.P.), DFG MA1081/19-1 (to D.J.M.) and R37HL036153 (to W.L.).

Author Contributions J.v.d.E. and S.R. designed the project. M.M. and D.J.M. provided protein samples. W.L. and D.J.M. provided information on tropomyosin and actin. J.v.d.E. performed research, analysed the data, and prepared figures. J.v.d.E. and P.A.P. improved image processing of helical specimens. S.R. managed the project and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The coordinates for the electron microscope structure has been deposited in the Electron Microscopy Data Bank under accession code EMD-6124. Coordinates of F-actin in complex with tropomyosin have been deposited in the Protein Data Bank with accession number 3J8A. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.R. (stefan.raunser@mpi-dortmund.mpg.de).

METHODS

Protein expression and purification. F-actin was isolated from rabbit skeletal muscle acetone powder and purified with several polymerization and depolymerization steps as described previously[29]. The composition of the G-actin buffer was 5 mM Tris pH 7.5, 1 mM DTT, 0.2 mM $CaCl_2$. Recombinant α-tropomyosin (*Mus musculus*, alpha-1 chain, Uniprot-ID P58771) with an alanine-serine extension was purified from *Escherichia coli*, based on the protocol of ref. 30. The α-tropomyosin isoform we used is abundant in smooth and striated muscles. The actin–tropomyosin complex was prepared by mixing F-actin with tropomyosin initially at a molar ratio of 7:1 in filament buffer (5 mM Tris-HCl pH 7.5, 1 mM DTT, 100 mM KCl, and 2 mM $MgCl_2$). The final concentration of tropomyosin for either negatively stained or frozen specimens was then adjusted empirically to obtain complete decoration and only few unbound tropomyosins in the background. For the cross-linking studies between F-actin and tropomyosin, samples were prepared in filament buffer containing 5 mM HEPES (pH 7.5).

Preparation for negative-stain electron microscopy. To stabilize and straighten the F-actin filaments (this is important for image processing), we decorated them with tropomyosin. Actin–tropomyosin complexes were directly prepared before the negative staining procedure. 4 μl of sample was applied on a freshly glow-discharged copper grid (Agar Scientific; G2400C) with an additional thin and continuous carbon film. After incubation for 60 s on the grid the sample was blotted using filter paper (Whatman no. 4), washed twice with filament buffer and stained with 0.75% uranyl formate (pH 4.0) for 30 s. For cross-linking studies, after incubation for 60 s of the sample solution on the grid and one washing step, a droplet of filament buffer with 0.25% glutardialdehyde was applied for 30 min on the grid before negative staining to avoid putative pH-induced artefacts. All images were taken with a JEOL JEM-1400 electron microscope equipped with a LaB_6 cathode at an operation voltage of 120 kV. Micrographs were recorded manually with a 4,000 × 4,000 pixel CMOS TemCam F416 (TVIPS) camera under minimal dose conditions.

Image processing of negatively stained filaments. All micrographs had a pixel size of 2.32 Å per pixel and filaments were boxed with a boxing distance of 13 pixels (overlap ~90%) and a total box size of 128 pixels. We analysed 8,371 segments from 40 images of bare F-actin, 27,926 segments from 111 images of F-actin decorated with tropomyosin, and 27,011 segments from 81 images of F-actin decorated with tropomyosin and cross-linked with glutardialdehyde. Subsequently, all segment stacks were aligned and classified independently using reference-free alignment and *k*-means classification procedures as implemented in SPARX[31] (Extended Data Fig. 1c–h, j). During the helical refinement and reconstruction using the *helicon* package implemented in SPARX[31], a cylinder filled with Gaussian noise was used as an initial three-dimensional template. The position of tropomyosin was determined in the resulting three-dimensional maps. The difference maps of tropomyosin–bare F-actin and glutardialdehyde–bare F-actin were calculated with the 'vop subtract' command in Chimera[32] (Extended Data Fig. 8c–g).

Grid preparation and image acquisition for cryo-EM. A 1.5-μl sample was applied to a glow-discharged holey carbon grid (C-flats 2/1, Protochips), incubated for 10 s and manually blotted for 3 s from the backside with filter paper (Whatman no. 5), before vitrification by plunging the grid into liquid ethane using a cryo plunger Cp3 (Gatan). Screening for the best sample and blotting conditions was performed on a JEOL JEM 3200FSC electron microscope equipped with a field emission gun and operated at a voltage of 200 kV. The omega in-column energy filter of the microscope was used to estimate the best ice conditions (~40–60 nm thickness). Finally, a data set was taken with a spherical-aberration corrected FEI Titan Krios transmission electron microscope equipped with an extra-high brightness field emission gun (XFEG) and operated at a voltage of 300 kV. Images were recorded with a back-thinned 4k × 4k FEI Falcon2 direct detection camera under minimal dose conditions using the automatic data collection software EPU (FEI). Within each selected grid-hole, three different positions were imaged, each with a total exposure of 1 s and a frame time of 55 ms. Seven frames from 85 ms to 475 ms with a total dose of 14.6 electrons per $Å^2$ and one total average with an electron dose of 30.7 electrons per $Å^2$ were used for image processing. The defocus range of the data set was 0.8–2.6 μm. The pixel size was 1.14 Å per pixel, based on the reflection of gold at the same magnification. Since the inaccuracy of pixel size determination is about 2% with this method, we used the well-defined rise per subunit of F-actin (27.6 Å) from fibre diffraction studies[14,33] instead to calibrate the pixel size, which was then 1.12 Å per pixel.

Image processing of the cryo-EM data set. As a first quality control, all 1,311 frame averages were manually inspected and only 689 images with high filament and ice quality (Extended Data Fig. 1a) were used for further processing. Frames were aligned and afterwards summed up using motion correction[34]. Filaments were manually selected with *helixboxer* in SPARX. A total of 109,242 segments from 7,854 filaments were extracted with a box size of 256 pixels and a boxing distance of 29 pixels (overlap ~90%). Thus, the approximate distance between them (~32 Å) slightly exceeded the rise of the helical assembly of actin (~27–28 Å), as required by the *helicon* design. The defocus and astigmatism and the accuracy of both were determined by *cter*[35] in SPARX. All segments were aligned and classified using reference-free alignment and *k*-means classification procedures implemented in SPARX (Extended Data Fig. 1b, i). For the initial reconstruction, the data set was binned twice to a pixel size of 2.24 Å and a box size of 128 pixels. We used a cylinder filled with Gaussian noise as a three-dimensional initial template for the helical refinement and reconstruction in *helicon*. After adjusting the pixel size as described above, the helical symmetry parameters converged to a rise per subunit of 27.5 Å and an azimuthal rotation of 166.4°. Initially, the entire F-actin–tropomyosin map was used as the reference during the first rounds of refinement, using the helical symmetry parameters of F-actin. However, tropomyosin spans over seven actin subunits and therefore has a symmetry different to that of actin. Nevertheless, because it is divided into seven pseudo-repeating units, each of which binds to a successive actin subunit along F-actin, the symmetry of F-actin could be used for reconstructions of actin-tropomyosin complexes of up to 6 Å resolution, accepting that the overlap region of the N and C termini of tropomyosin would not be discernible in these maps. Once the refined map reached a resolution of 6.5 Å, as determined by the $FSC_{0.5}$ criterion[36], we masked out tropomyosin in the reference volume before continuing with high-resolution refinements of F-actin. In the last refinement iteration, we used only filaments derived from summed up frames with a drift of less than 7 Å, as estimated in the motion correction[34] step and with a sufficient content of high-resolution information, as determined by *cter* in SPARX. This yielded a subset of 74,228 segments for the final map, which corresponded to ~120,000 asymmetric subunits. The segments' statistics are given in Extended Data Table 1. Since an electron microscope map of a filament is truncated at the edges of the cubic volume of the box (here 256 voxels), a mask with Gaussian edges must be applied for resolution estimation to minimize artefacts. Therefore, Fourier shell correlation (FSC) analysis was performed within the central area of the volume, resulting in a nominal resolution of 3.7 Å ($FSC_{0.5}$ criterion[36]) for the F-actin cryo-EM density map (Extended Data Fig. 2a). The map of F-actin was then sharpened using a negative *b* factor of –50 $Å^2$ and filtered to its nominal resolution. Finally, the tropomyosin map filtered to 6.5 Å was merged with the final F-actin map to obtain a map of the entire F-actin–tropomyosin complex.

We also tried to process tropomyosin separately from F-actin. After alignment of the filaments based on F-actin, we masked out tropomyosin. To find the N- or C-terminal overlap and guarantee the proper alignment of tropomyosin, we used very large boxes to allow tropomyosin to shift by almost its full length and tried to align the molecules. However, this approach did not work. We suspect that the pseudo-repeats and even the N/C-terminal overlaps are too similar to each other (at least in cryo-EM raw images), resulting in an improper alignment. We are therefore investigating labelling strategies to solve this problem.

Model building and refinement of F-actin. The central F-actin subunit (chain A) shows all available contacts to adjacent chains (B, C, D, E) and was therefore used for further structural analysis. In a first step, five central subunits (chains A–F) of an F-actin model (PDB accession code 4A7N; ref. 8) were rigid-body fitted into the electron microscope map, using 'Fit in Map' and the map was initially segmented with 'Split Map' in Chimera[37]. Homology modelling was performed using MODELLER[38] and the resulting model of the central subunit was then flexibly fitted into its corresponding part of the density using DireX[39]. Finally, the flexibly fitted subunit was used to create a starting F-actin model of five adjacent subunits (hereafter referred to as the 'pentamer') for further model building. The cryo-EM density was converted to structure factors with the CCP4 program suite[40]. The model and the map were then used for real space refinement and model building in COOT[41]. After a single iteration of model building of the central F-actin subunit, its part of the map was further used for an initial automatic reciprocal space refinement using PHENIX.refine[42]. Since this part was less well resolved in the N and C terminus, and model building was difficult for a part of the D-loop, we deleted amino acids 1–4, 42–50 and 372–375 from the model. Owing to the high reliability of the experimental phases of the cryo-EM map, we were able to apply experimental phase restraints during refinement in PHENIX. After refinement the output model was again manually refined in COOT using the original cryo-EM map to minimize model bias. Rebuilding of the N or C terminus was not possible even in the refined map (Extended Data Fig. 2c, d). However, after refinement in PHENIX the definition of the D-loop improved, so ultimately it could be manually rebuilt *ab initio*.

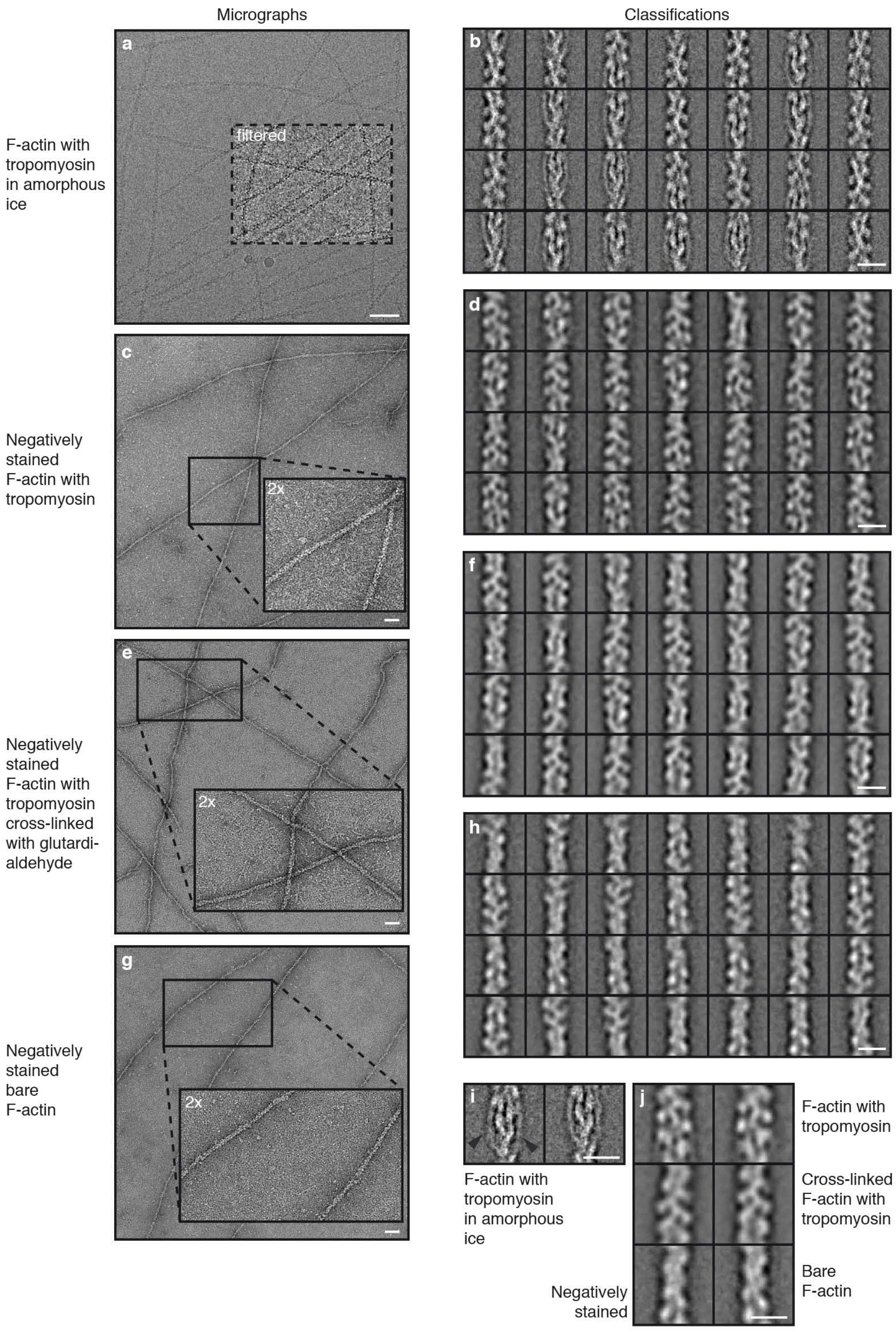
After refinement of the subunit had converged and model building was completed, the resulting subunit was used to assemble the F-actin pentamer and the respective full map of the central pentamer was then segmented. The molecular model and the respective original density were then used as input for PHENIX.refine. Again, we alternated manual model building and refinement steps. Finally, the resulting model was validated using MOLPROBITY[43] (see data statistics in Extended Data Table 1). For further structural analysis we used the final model and the cryo-EM map of the actin pentamer.

Structure analysis and visualization. To describe the interaction between F-actin and tropomyosin, a chimaeric molecular model of tropomyosin[44] was initially rigid-body fitted into the cryo-EM density of F-actin–tropomyosin. Subsequently, the respective density of tropomyosin was extracted as described above using Chimera. The molecular model of tropomyosin was then flexibly fitted into the resulting density using DireX. Furthermore, during flexible fitting, in order to preserve its electrostatically preferred position on F-actin we fixed tropomyosin in its vertical position on the filament. Atomic clashes and stereochemistry were repaired using PHENIX.geometry_minimization[42]. Since the tropomyosin N/C-terminal overlap is not fully understood[45], we reduced the model to the five central pseudo-repeats. Owing to the limited resolution of the cryo-EM density in the region of tropomyosin, we avoided interpretation of tropomyosin on the single-amino-acid level.

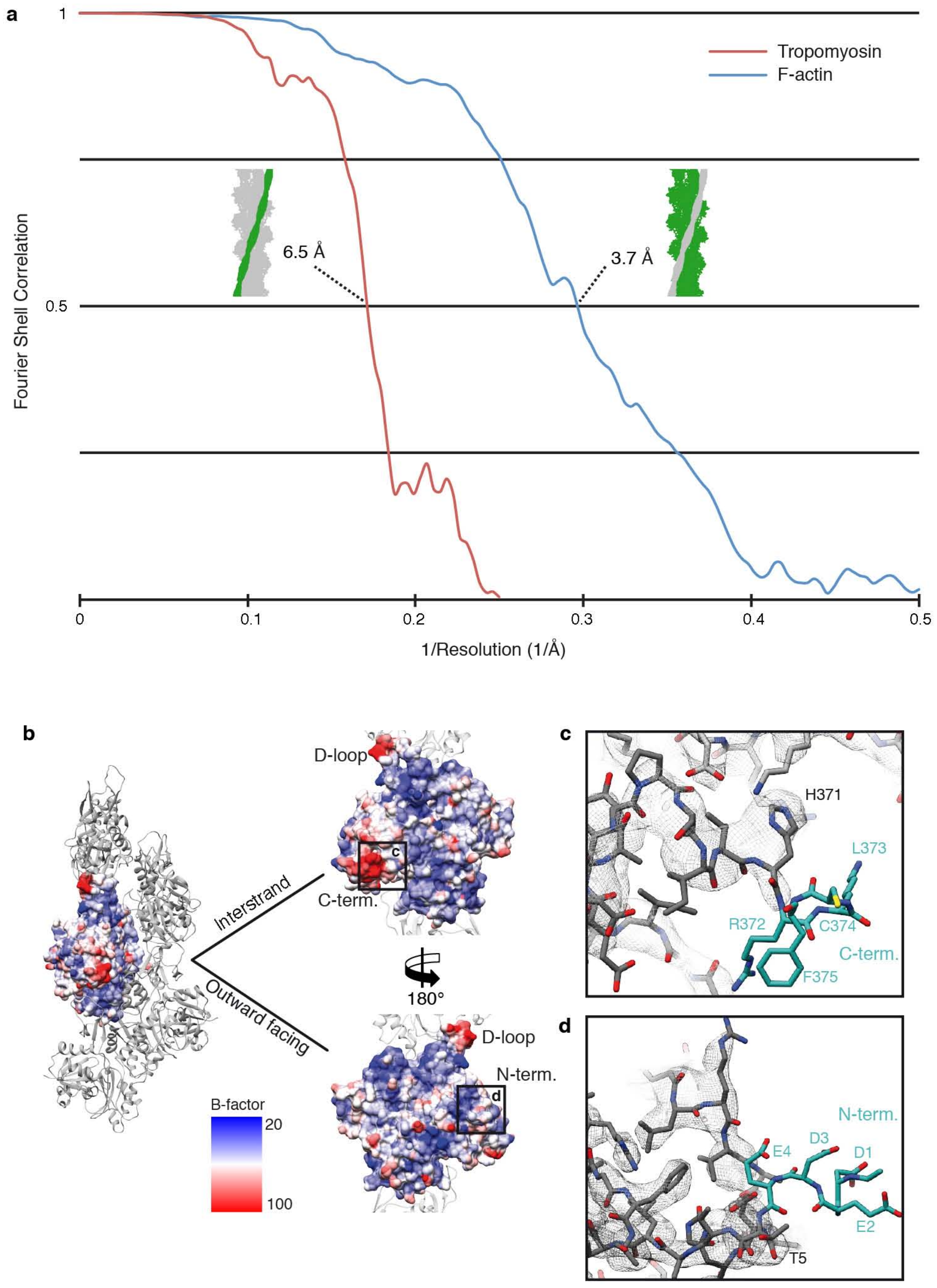
We used the refined actin pentamer and the modelled tropomyosin to reconstruct a full filament for visualization of the actin–tropomyosin interaction (Fig. 4). The filament was protonated using H++[46] at different pH values and the electrostatic coulomb potential of the filament surface was calculated to range from -10 kcal mol^{-1} to $+10$ kcal mol^{-1} in Chimera (Fig. 4b, Extended Data Fig. 8h). For visualization of the hydrophobicity per amino acid residue, we used 'Define attribute' in Chimera and generated amino-acid-specific scores[47]. Point mutations were introduced via COOT or Chimera to show charge-induced repulsions and changes in hydrophobicity. The HGMD[48] library was browsed to find mutations in regions of interest.

For detailed analyses of intra- and intermolecular interactions in F-actin, we applied PISA tools[49] and distance measurements in Chimera. The G- to F-actin transition within a monomer/subunit was visualized and interpreted with the 'Morph Conformations' in Chimera. Furthermore, the hinge region and the global rotation angle were detected with the DynDom Sever[50] using an atomic model of G-actin (PDB accession code 1J6Z; ref. 51) and one subunit of our F-actin model (Fig. 3, Extended Data Fig. 7).

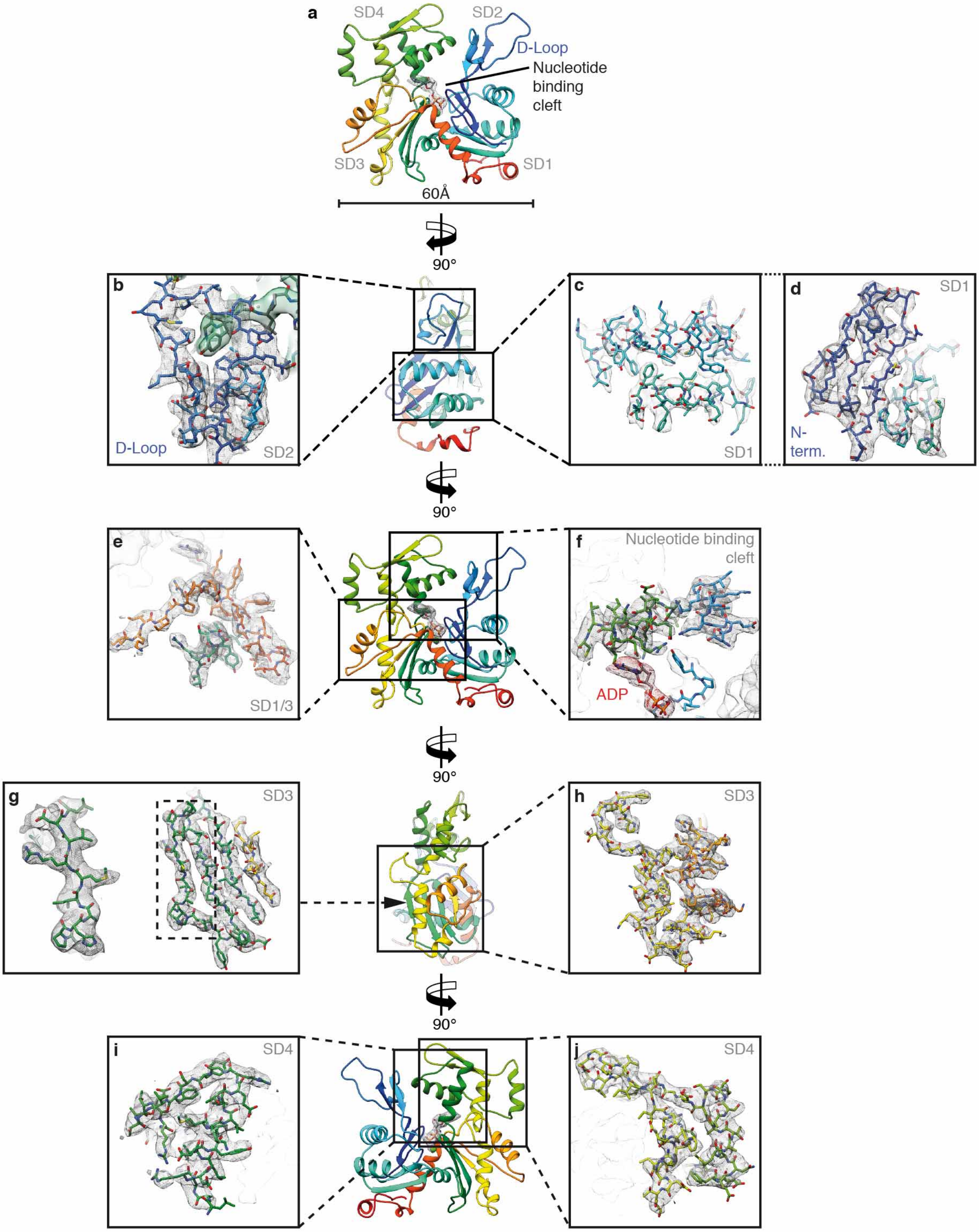
29. Pardee, J. D. & Spudich, J. A. Purification of muscle actin. *Methods Enzymol.* **85B,** 164–181 (1982).
30. Coulton, A., Lehrer, S. S. & Geeves, M. A. Functional homodimers and heterodimers of recombinant smooth muscle tropomyosin. *Biochemistry* **45,** 12853–12858 (2006).
31. Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157,** 47–55 (2007).
32. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
33. Dominguez, R. & Holmes, K. C. Actin structure and function. *Annu. Rev. Biophys.* **40,** 169–186 (2011).
34. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10,** 584–590 (2013).
35. Penczek, P. A. *et al.* CTER-rapid estimation of CTF parameters with error assessment. *Ultramicroscopy* **140,** 9–19 (2014).
36. Penczek, P. A. Resolution measures in molecular electron microscopy. *Methods Enzymol.* **482,** 73–100 (2010).
37. Pintilie, G. D., Zhang, J., Goddard, T. D., Chiu, W. & Gossard, D. C. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170,** 427–438 (2010).
38. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815 (1993).
39. Wang, Z. & Schröder, G. F. Real-space refinement with DireX: from global fitting to side-chain improvements. *Biopolymers* **97,** 687–697 (2012).
40. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242 (2011).
41. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).
42. Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* **55,** 94–106 (2011).
43. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 12–21 (2010).
44. Li, X. E., Orzechowski, M., Lehman, W. & Fischer, S. Structure and flexibility of the tropomyosin overlap junction. *Biochem. Biophys. Res. Commun.* **446,** 304–308 (2014).
45. Lehman, W., Li, X. E., Orzechowski, M. & Fischer, S. The structural dynamics of α-tropomyosin on F-actin shape the overlap complex between adjacent tropomyosin molecules. *Arch. Biochem. Biophys.* **552–553,** 68–73 (2014).
46. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **40,** W537–W541 (2012).
47. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433,** 377–381 (2005).
48. Stenson, P. D. *et al.* The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum. Genet.* **133,** 1–9 (2014).
49. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797 (2007).
50. Hayward, S. & Lee, R. A. Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J. Mol. Graph. Model.* **21,** 181–183 (2002).
51. Otterbein, L. R., Graceffa, P. & Dominguez, R. The crystal structure of uncomplexed actin in the ADP state. *Science* **293,** 708–711 (2001).
52. Bathe, F. S., Rommelaere, H. & Machesky, L. M. Phenotypes of myopathy-related actin mutants in differentiated C2C12 myotubes. *BMC Cell Biol.* **8,** 2 (2007).
53. Costa, C. F. *et al.* Myopathy mutations in alpha-skeletal-muscle actin cause a range of molecular defects. *J. Cell Sci.* **117,** 3367–3377 (2004).
54. Nowak, K. J. *et al.* Mutations in the skeletal muscle alpha-actin gene in patients with actin myopathy and nemaline myopathy. *Nature Genet.* **23,** 208–212 (1999).
55. Ohlsson, M., Tajsharghi, H., Darin, N., Kyllerman, M. & Oldfors, A. Follow-up of nemaline myopathy in two patients with novel mutations in the skeletal muscle alpha-actin gene (ACTA1). *Neuromuscul. Disord.* **14,** 471–475 (2004).
56. Ilkovski, B. *et al.* Evidence for a dominant-negative effect in ACTA1 nemaline myopathy caused by abnormal folding, aggregation and altered polymerization of mutant actin isoforms. *Hum. Mol. Genet.* **13,** 1727–1743 (2004).
57. Graziano, C., Bertini, E., Minetti, C. & Porfirio, B. Alpha-actin gene mutations and polymorphisms in Italian patients with nemaline myopathy. *Int. J. Mol. Med.* **13,** 805–809 (2004).
58. Perkins, K. Z. *et al.* P3.48 Exome sequencing with linkage analysis identifies a novel ACTA1 variant in a large family with progressive muscle weakness. *Neuromuscul. Disord.* **21,** 696–697 (2011).
59. Nair, U. B. *et al.* Crystal structures of monomeric actin bound to cytochalasin D. *J. Mol. Biol.* **384,** 848–864 (2008).
60. Sudo, A. *et al.* Sibling cases of severe infantile form of nemaline myopathy with ACTA1-gene mutation. [in Japanese] *No To Hattatsu* **45,** 452–456 (2013).
61. Yao, X., Grade, S., Wriggers, W. & Rubenstein, P. A. His(73), often methylated, is an important structural determinant for actin. A mutagenic analysis of HIS(73) of yeast actin. *J. Biol. Chem.* **274,** 37443–37449 (1999).
62. Agrawal, P. B. *et al.* Heterogeneity of nemaline myopathy cases with skeletal muscle alpha-actin gene mutations. *Ann. Neurol.* **56,** 86–96 (2004).
63. D'Amico, A. *et al.* Fatal hypertrophic cardiomyopathy and nemaline myopathy associated with ACTA1 K336E mutation. *Neuromuscul. Disord.* **16,** 548–552 (2006).
64. Stenzel, W. *et al.* Fetal akinesia caused by a novel actin filament aggregate myopathy skeletal muscle actin gene (ACTA1) mutation. *Neuromuscul. Disord.* **20,** 531–533 (2010).
65. Jain, R. K. *et al.* Nemaline myopathy with stiffness and hypertonia associated with an ACTA1 mutation. *Neurology* **78,** 1100–1103 (2012).
66. Kollmar, M., Dürrwang, U., Kliche, W., Manstein, D. J. & Kull, F. J. Crystal structure of the motor domain of a class-I myosin. *EMBO J.* **21,** 2517–2525 (2002).



Extended Data Figure 1 | Micrographs and classifications of different data sets. a–h, Representative digital micrographs and corresponding representative two-dimensional class averages of F-actin decorated with tropomyosin in amorphous ice (in total 300 class averages of randomly chosen 40,000 phase-flipped segments of in total 109,242 segments from 689 images) (**a**, **b**), negatively stained (in total 300 class averages of 27,926 segments from 111 images) (**c**, **d**), negatively stained after cross-linking with glutardialdehyde (in total 300 class averages of 27,011 segments from 81 images) (**e**, **f**), and a micrograph of negatively stained bare F-actin (in total 100 class averages of 8,371 segments from 40 images) (**g**, **h**). Scale bars, 50 nm. Each class average ('Classifications') contains 130–200 (cryo data set) or 70–90 (negatively stained data sets) single segments. Scale bars, 10 nm. A boxed region of the digital micrograph in **a** was band-pass filtered to allow a better visualization of the filaments. Insets in **c**, **e** and **g** show 2× magnified regions of the digital micrographs. **i**, Two of the class averages depicted in **b** that show a clear tropomyosin density. The tropomyosin density is indicated by black arrow heads. Scale bar, 10 nm. **j**, Comparison of representative class averages of the three negative-stain data sets. Class averages of the data sets with tropomyosin (top panels) show additional density and a larger diameter than bare F-actin (bottom panel). Scale bar, 10 nm.



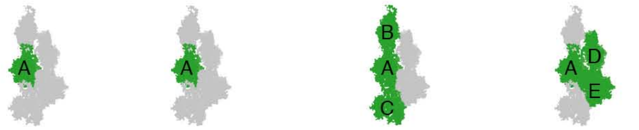
Extended Data Figure 2 | Resolution of the F-actin–tropomyosin complex. **a**, FSC curves of different areas of interest (green) by masking (see Methods). The resolution of tropomyosin was estimated using the twice down-sampled data set (2.24 Å per pixel). The $FSC_{0.5}$ criterion indicates that the tropomyosin density map has a resolution of 6.5 Å. The resolution of the final F-actin density map is estimated at a resolution of 3.7 Å. **b**, Surfaces of F-actin with B-factors (high is red, low is blue) estimated by the reciprocal space refinement in PHENIX[42]. A side view as well as top views on the inward- and outward-facing surfaces, that is, facing the adjacent F-actin subunit inside the filament and oriented to the periphery, respectively, are shown. **c**, **d**, Putative structures of the C and N termini (cyan), respectively. Map density is missing in these regions.



Extended Data Figure 3 | Representative regions of the F-actin cryo-EM map. a, Overview of the atomic model of an F-actin subunit rainbow-coloured from the N terminus (blue) to the C terminus (red). ADP and the coordinated cation, probably Mg^{2+}, are depicted inside their corresponding map densities. **b–d**, Side views. Interface between the D-loop (SD2) and the SD3 of the adjacent actin subunit (**b**), outer two helices of SD1 (**c**), and inner β-sheet with N terminus (**d**). **e**, **f**, Front views. Connection of SD1 and SD3 (**e**) and nucleotide binding cleft (**f**). **g**, **h**, Side views. Inner β-sheet of SD3 with one highlighted strand (**g**) and outer part of SD3 (**h**). **i**, **j**, Back views. SD4 separated in two parts.

a

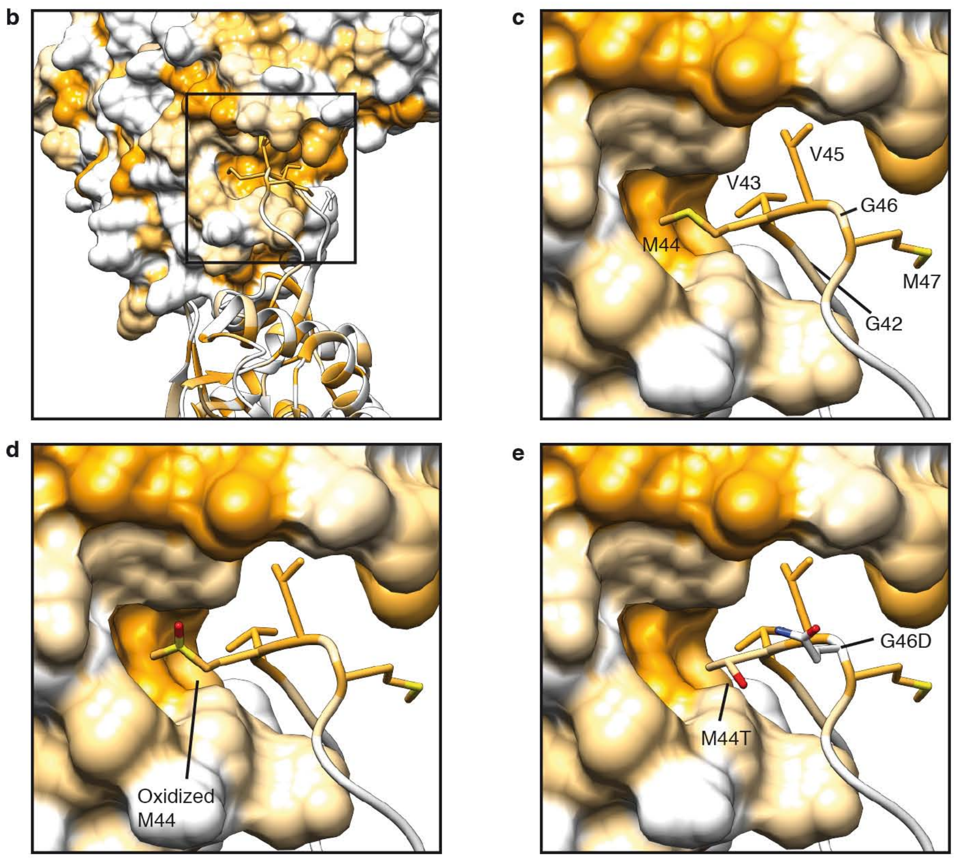
Filament stability by intra/intermolecular interactions



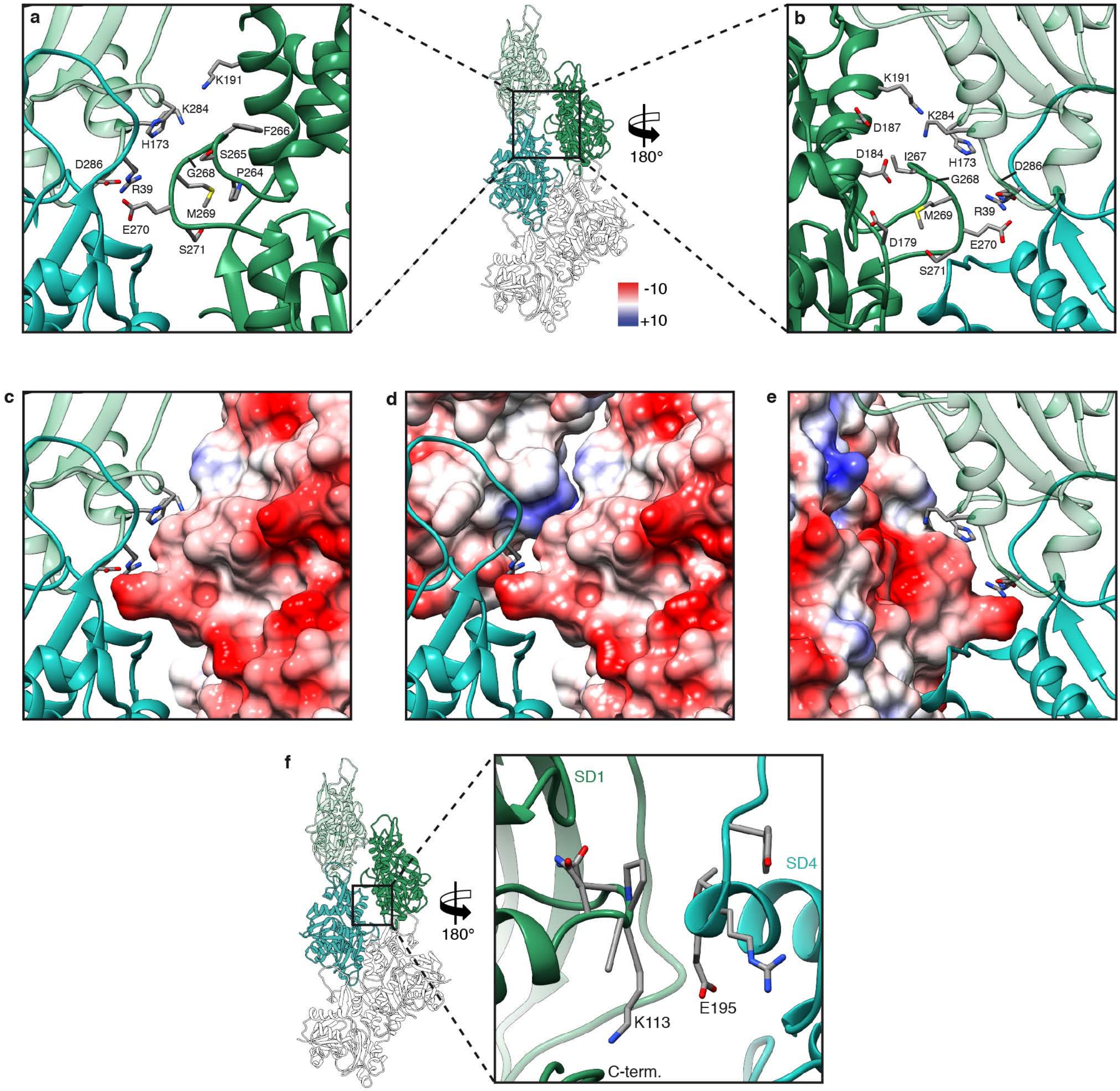
Chain A	A	B	C	D	E	Function	Mutations*
Arg37	Asp80, Asp81		Same as chain B with A		Same as chain D with A	D-Loop stabilization	
Arg39		Asp286		Glu270		F-Actin/D-Loop stabilization	D286G[52-54]; E270Q[15]
Asp51	Arg37, Lys84					D-Loop stabilization	
Tyr53	Lys61	Glu167				F-Actin/D-Loop stabilization	
Arg62		Asp288				F-Actin/D-Loop stabilization	D288N[15]
Tyr69	Arg183					D-Loop stabilization	
His173				Ile267, Gly268		F-Actin stabilization	G268R[3,52-53], G268D[55], G268C[53-56], G268S[57]
Glu195				Lys113		F-Actin stabilization	K113E[15]; E195D†[58]
Asp244		Arg290				F-Actin stabilization	
Gly42-Gly48		SD3				F-Actin/D-Loop stabilization	oxidized M44, M47V[20];
Pro38-Lys50		Tyr169					M44T, G46D, G46C, M47V[15]

* Mutations connected to nemaline myopathies if not labeled differently

† Mutation connected to neuromuscular disorders



Extended Data Figure 4 | Overview of inter- and intrastrand interactions and hydrophobic D-loop. **a**, Table of identified residues that are involved in intra- and intermolecular interactions and known mutations. Mutation data are from refs 3, 15, 20 and 52–58. **b**, Overview of the D-loop bound to the hydrophobic cleft in SD3 of the F-actin molecule on top. Surfaces and residues are coloured from high (yellow) to low (white) hydrophobicity. **c**, Involved hydrophobic residues of the D-loop. **d**, **e**, Modifications in the D-loop, like oxidation of methionine 44 (ref. 20) (**d**) or mutations of methionine 44 and glycine 46 (M44T, G46G)[15,20] (**e**) change the polarity or insert charges and consequently weaken the hydrophobic interactions. Thus, this destabilization of the intrastrand contact is connected to anomalous actin filament assembly and nemaline myopathies.



Extended Data Figure 5 | Inter- and intrastrand F-actin interactions. **a**–**e**, The interface at the plug involves three residues (R39, E270, D286) that form salt bridges and mediate not only one interstrand contact but also one intrastrand contact (see also Fig. 2a). In addition, the orientation of residues 264–269 result in a negatively charged patch that electrostatically interacts with positively charged residues on the opposing actin. **a**, **b**, Front and back view of the interface at the plug, respectively. **c**–**e**, Surface representations (front views in **c** and **d**, back view in **e**) depicting the Coulomb potential (ranging from $-10\,\mathrm{kcal\,mol^{-1}}$ (red) to $+10\,\mathrm{kcal\,mol^{-1}}$ (blue) at pH 7.5), indicating that the interaction of the upper region of the plug with adjacent inter- or intrastrand molecules is mediated by electrostatic interactions. **f**, Another interstrand contact is formed by residues 110–115 of SD1 and residues 191–199 of SD4 of the adjacent actin. However, no prominent electrostatic or hydrophobic interactions could be identified at this interface.

a

Intramolecular interaction by salt bridges and ADP/Cation coordination in the nucleotide binding cleft

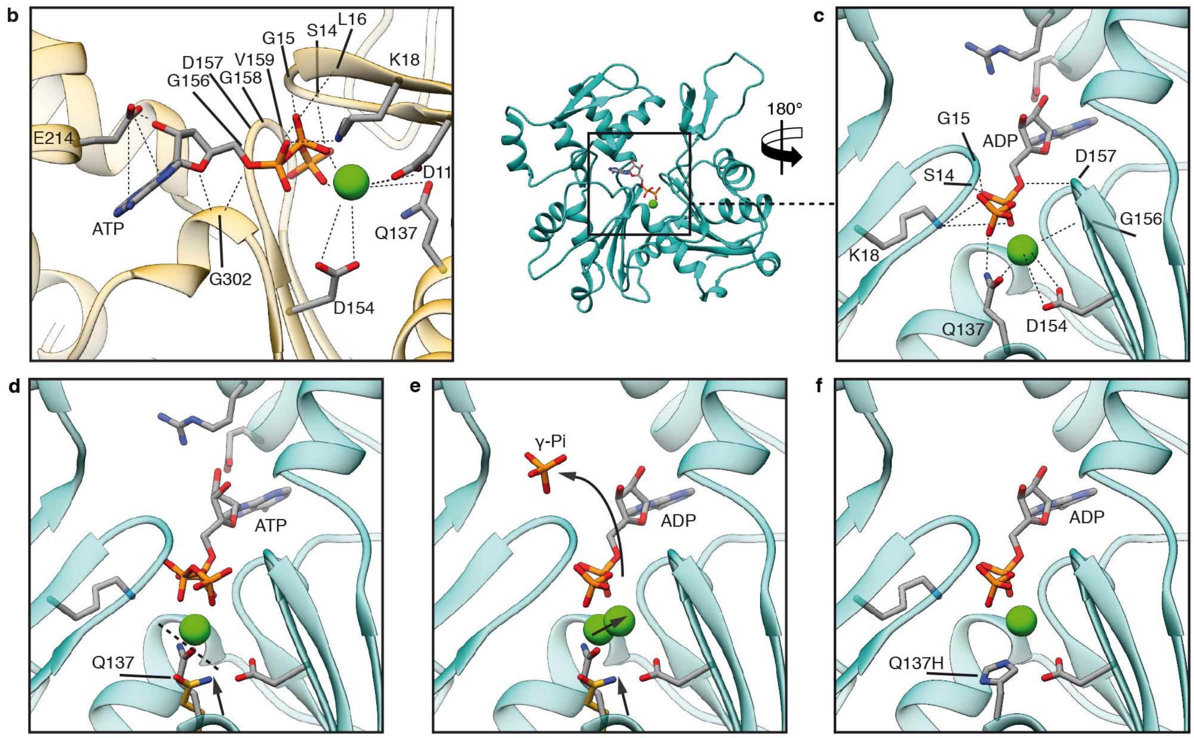
Chain A	A	Function	Mutations*
Tyr69	Arg183	Cleft stabilization	R183C[53-54,60]; R183G[53,56], R183L[15]; R183S[3]
His73	Asp179	Cleft stabilization	H73N[15]; H73L, H73R, D179N, D179H[3]; H73R[61]; D179G[62]
Asp157	Arg183	Cleft stabilization	R183C[53-54,60]; R183G[53,56], R183L[15]; R183S[3]
Asp187	Arg206	Cleft stabilization	
Asp207	Arg210	Cleft stabilization	E207D[57]
Asp211	Arg210	Cleft stabilization	
Lys213	Tyr306	Cleft stabilization	
Tyr337	Lys18, Lys336	Cleft stabilization	K336E†[63]; K336I[3]
Ser14, Gly15	ADP	ADP binding	G15R‡[52-54], G15D§[64]
Lys18	ADP	ADP binding	
Gln137	ADP	ADP binding	Q137H[23]
Gly156, Asp157	ADP	ADP binding	
Arg210	ADP	ADP binding	
Glu214	ADP	ADP binding	
Gly302	ADP	ADP binding	
Gln137	Ion	Ion coordination	Q137H[23]
Asp154	Ion	Ion coordination	
Gly156	Ion	Ion coordination	

* Mutations connected to nemaline myopathies if not labeled differently

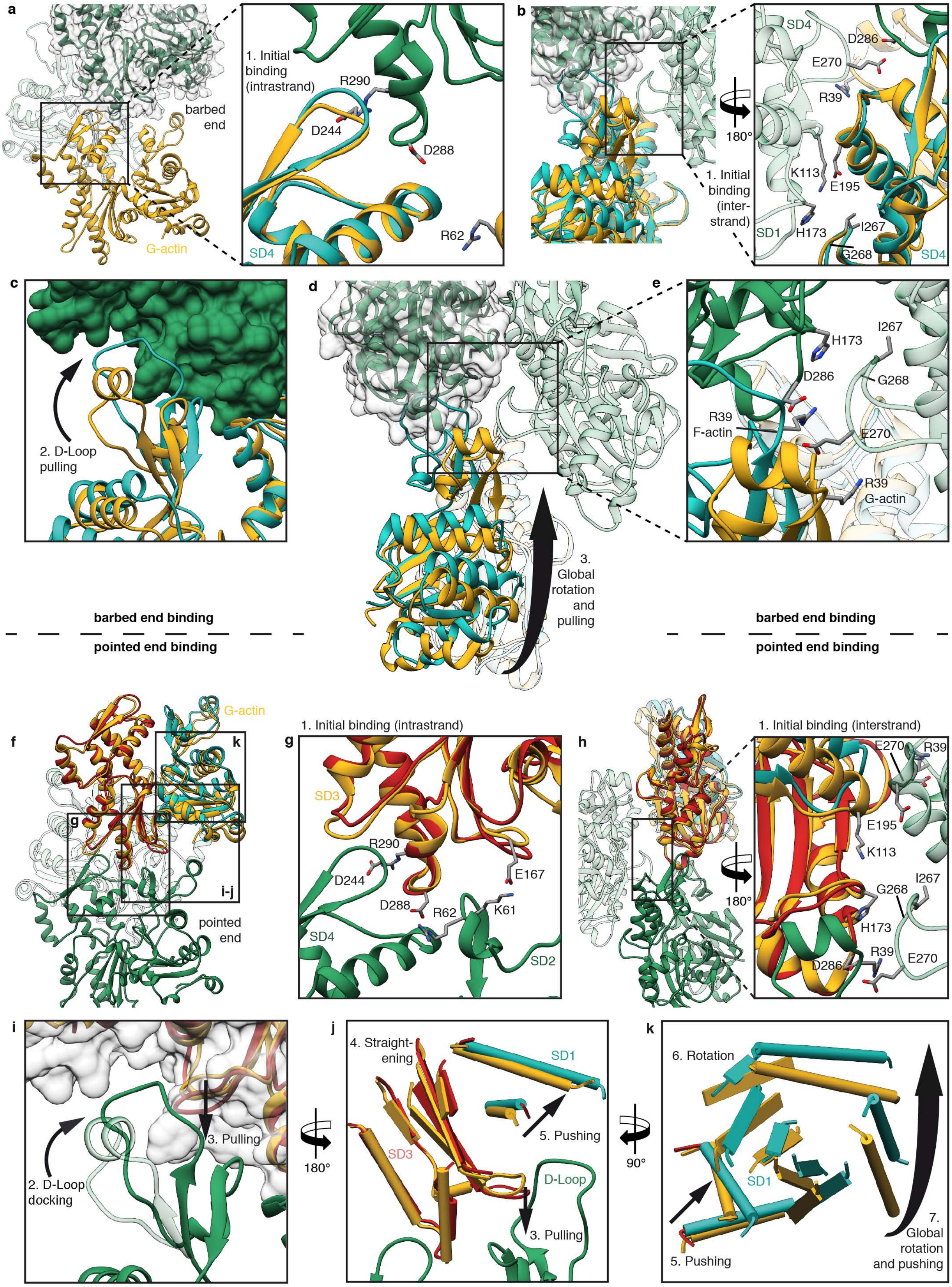
† Mutation connected to nemaline myopathies and hypertrophic cardiomyopathies

‡ Mutation connected to actin myopathies

§ Mutation connected to fetal akinesia



Extended Data Figure 6 | Nucleotide binding site and intramolecular interactions. **a**, Table of identified intramolecular interactions that result in a stabilization of the nucleotide binding cleft, coordination of ADP and a divalent cation. Mutation data are from refs 3, 15, 23, 53, 54, 56, 57 and 60–64. **b**, Coordination of ADP and Ca^{2+} in the nucleotide binding cleft in G-actin (PDB accession code 3EL2; ref. 59). **c**–**e**, Back views of the nucleotide binding cleft of F-actin (cyan) with bound ADP–Mg^{2+} or ADP–Ca^{2+} (**c**), ATP–Ca^{2+} (**d**, relative position taken from PDB accession code 3EL2; ref. 59) and conformational changes between the G-actin-ATP and F-actin-ADP state (**e**). Glutamine 137 is moved closer to ADP, coordinating not only the cation (as in G-actin) but also the nucleotide β-phosphate (**c**). The presence of ATP instead of ADP in the nucleotide-binding site would be sterically unfavourable, suggesting that a different intermediate conformation exists for F-actin-ATP (**d**). The shorter distance of glutamine 137 to the γ-phosphate probably induces ATP hydrolysis and then afterwards the cation takes the position of the γ-phosphate in the ADP-state (**e**). For comparison the position of glutamine 137 in G-actin is shown in yellow and the transition from G-actin to F-actin is depicted by arrows. **f**, Mutation of glutamine 137 to histidine results in hampered coordination of the ion and the nucleotide and is connected to nemaline myopathies[23].


a
1. Initial binding (intrastrand)
R290
D244
D288
R62
SD4
barbed end
G-actin
b
SD4
D286
E270
R39
180°
K113
E195
1. Initial binding (inter-strand)
SD1
H173
I267
G268
SD4
c
2. D-Loop pulling
d
3. Global rotation and pulling
e
H173
I267
D286
G268
R39 F-actin
E270
R39 G-actin
barbed end binding
pointed end binding
barbed end binding
pointed end binding
f
G-actin
k
g
i-j
pointed end
1. Initial binding (intrastrand)
SD3
R290
D244
E167
D288
R62
K61
SD4
SD2
h
1. Initial binding (interstrand)
E270
R39
E195
K113
180°
I267
G268
H173
R39
E270
D286
i
2. D-Loop docking
3. Pulling
180°
j
4. Straight-ening
SD1
5. Pushing
SD3
D-Loop
3. Pulling
90°
k
6. Rotation
SD1
5. Pushing
7. Global rotation and pushing

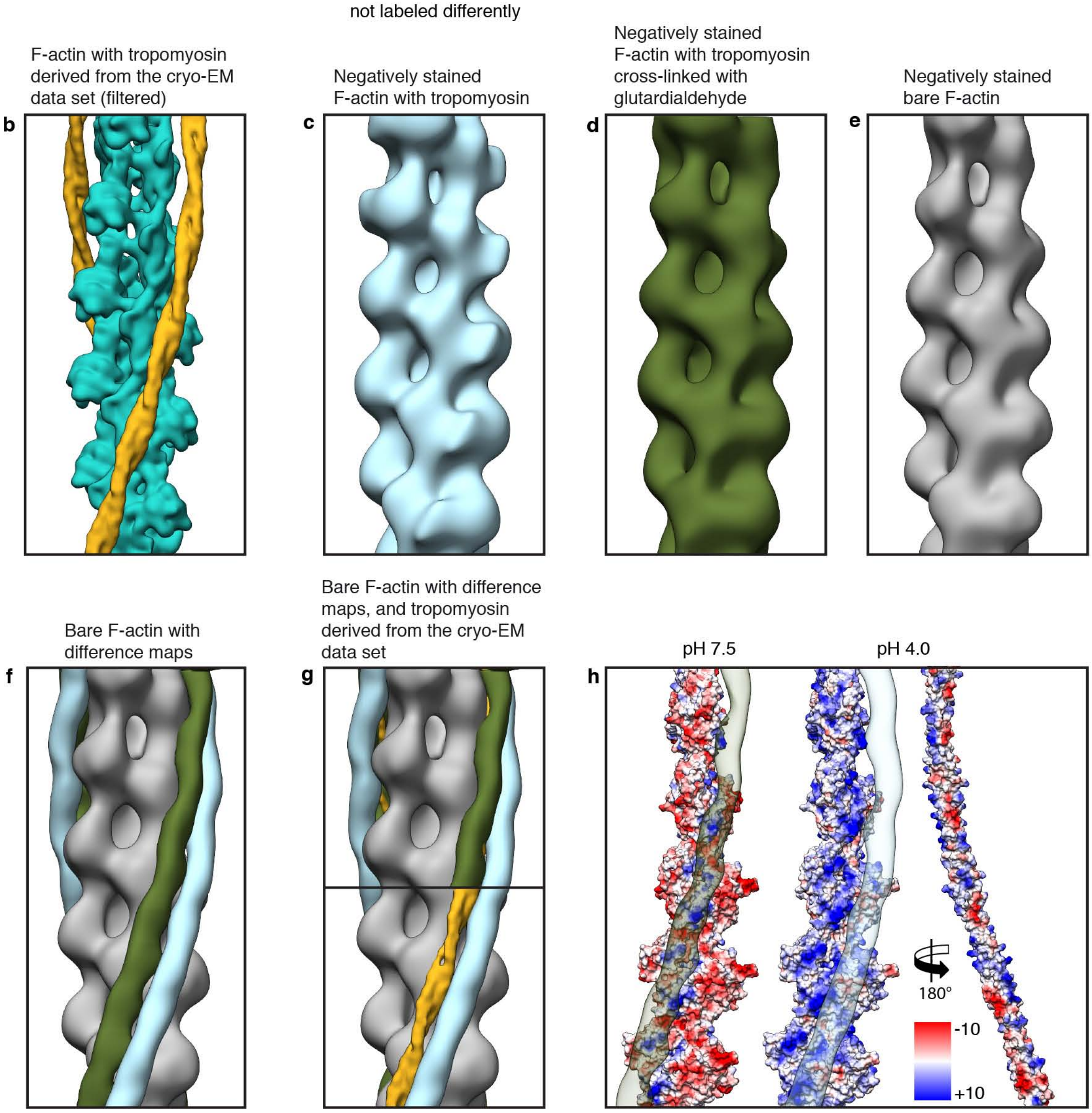
Extended Data Figure 7 | Model of barbed-end and pointed-end binding and G- to F-actin transition based on a comparison of start point (G-actin) and end point (F-actin). **a**, **b**, Binding of new G-actin (yellow, PDB accession code 3EL2; ref. 59) at the barbed end of F-actin (green) is initiated by the intrastrand binding of SD4 of G-actin to SD3 of F-actin (**a**, side view) and the interstrand binding of SD4 and SD1 (**b**, front view), respectively. G-actin is overlaid with a subunit of the structure found in F-actin (cyan). The main interstrand contacts are already present at the start of the transition and thereby guide the binding G-actin to its correct position, determining the symmetry of the filament. **c**, After initial binding the D-loop is trapped in the hydrophobic cleft of SD3 of F-actin and pulls on SD2. F-actin is depicted in surface representation (green). **d**, Finally, the SD2 and concomitantly SD1 are rotated and the final F-actin conformation is stabilized by various intramolecular interactions (see Figs 2c–f and 3c–f, Extended Data Fig. 6a). **e**, Owing to the relatively large distance from the SD1 and SD2 subdomains of the newly bound G-actin to the subunit of the opposite strand, the only influence of the opposite strand on the binding of G-actin is at the docking position of the D-loop. The transition of R39 illustrates well the docking of the D-loop. The residue is depicted in both states (G-actin and F-actin). **f**–**h**, Binding of new G-actin (yellow, PDB accession code 3EL2; ref. 59) at the pointed end of F-actin (green) is initiated by initial intrastrand binding (**g**) of SD3 of G-actin to SD4 of F-actin at the pointed end and interstrand binding of SD4 to SD3 (**h**), respectively. Again, the main interstrand contacts are available before the transition of G-actin (yellow) to F-actin (SD1 and SD2 in cyan, SD3 and SD4 in red). **i**, During binding to F-actin the transition from G-actin to F-actin is initiated by an induced fit of the F-actin D-loop to the hydrophobic cleft of the newly bound G-actin. This leads to a pulling down of the central β-sheet of SD3 of G-actin. **j**, The β-sheet is thereby straightened and pushes up two adjacent helices of SD1. **k**, The slight dislocation of these helices is transmitted to other regions of SD1 and thereby emphasized. This leads to a global rotation of SD1, which results in a considerable rotation of SD2 by an angle of 20° and a closure of the nucleotide binding cleft (Fig. 3; see Supplementary Video 3).

a

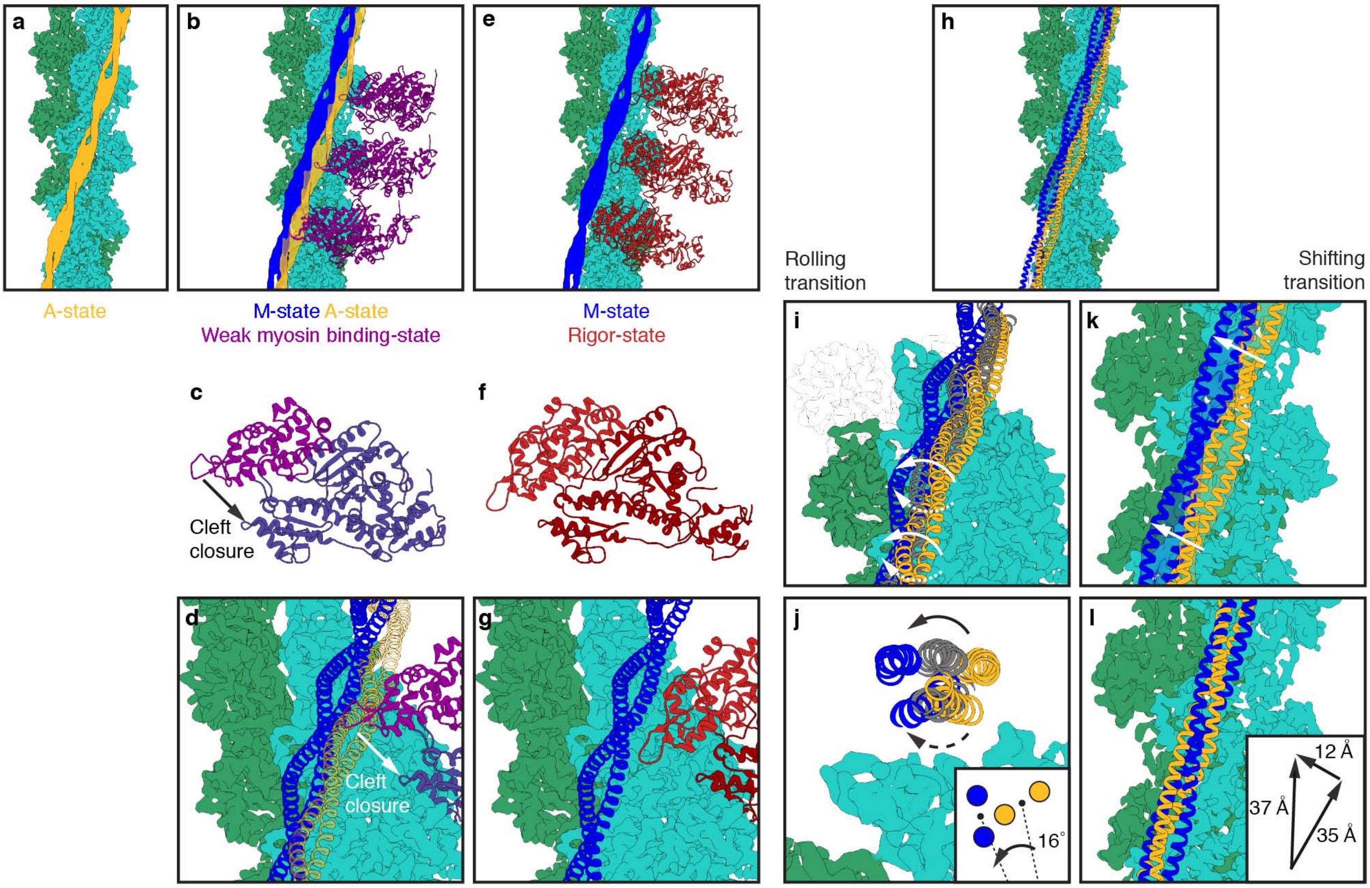
Tropomyosin binding to F-actin in A-state

Actin	Mutations*
Arg147	R147K[15]
Lys215	
Lys291	
Asp311	
Lys326	K326N[15,65]
Lys328	

* Mutations connected to nemaline myopathies if not labeled differently



Extended Data Figure 8 | Tropomyosin binding and comparison of reconstructions regarding the tropomyosin position on F-actin. **a**, Table showing putative residues of F-actin involved in tropomyosin binding and known mutations. Mutation data are from refs 15 and 65. **b**–**e**, Reconstructions of F-actin decorated with tropomyosin calculated from: the cryo-EM data set filtered to 15 Å (**b**), from a negatively stained data set (**c**), from a negatively stained data set after cross-linking with glutardialdehyde (**d**), and from negatively stained bare F-actin (**e**). **f**, By calculating a difference map between tropomyosin–bare F-actin (blue) and glutardialdehyde–bare F-actin (green), differences in the tropomyosin position on bare F-actin (grey) are visualized. **g**, Overlay of difference maps showing that the position of tropomyosin in the cryo-EM reconstruction (yellow) corresponds to the tropomyosin position of the negatively stained data set with the cross-linked complex (green). **h**, Surface of F-actin and tropomyosin (pseudo-repeats 2–6) with the electrostatic Coulomb potentials at pH 7.5 and pH 4, ranging from $-10\,\mathrm{kcal\,mol^{-1}}$ (red) to $+10\,\mathrm{kcal\,mol^{-1}}$ (blue) (see also Fig. 4b). Tropomyosin was rotated by 180° and shifted to the right to allow a better view of the F-actin–tropomyosin interface. Difference maps of the glutardialdehyde–bare F-actin and the tropomyosin–bare F-actin map are shown on the F-actin surface at pH 7.5 (left) and pH 4.0 (middle), respectively.



Extended Data Figure 9 | Model of tropomyosin transition on F-actin during myosin binding. **a**, Cryo-EM structure of the F-actin–tropomyosin complex with tropomyosin in the A-state. Tropomyosin (yellow, A-state), F-actin (green). **b**–**d**, Initial weak binding of myosin (magenta, PDB accession code 1LKX; ref. 66) to the F-actin–tropomyosin filament in the absence of troponin. Most of the myosin binding sites on F-actin are not occupied by tropomyosin and only loop 4 and the cardiomyopathy loop are sterically hindered from binding to the F-actin filament (**b**). Actin-induced closure of the 50-kDa cleft of myosin (**c**) results in a strong binding of myosin and tropomyosin moves to its M-state position (blue) (**d**). **e**–**g**, F-actin–tropomyosin–myosin complex in the rigor state (PDB accession code 4A7H; ref. 8). Myosin is shown in red. **h**–**l**, There are two possible ways for the transition of tropomyosin from the A-state to the M-state. Tropomyosin either rolls (**i**, **j**) or slides (**k**, **l**) from one to the other position. Rolling would involve an azimuthal rotation of ~16° with respect to the F-actin axis (inset of **j**) and a left-handed rotation of ~70° (indicated by solid arrows) or a right-handed rotation of ~110° (indicated by dotted arrows) with respect to its own axis (**i**, **j**). Sliding would imply not only an azimuthal shift of ~12 Å (indicated by white arrows in **k**), but also a tremendous shift of a half-tropomyosin repeat (that is, ~35 Å) along the F-actin filament (**k**, **l**). The radius to the filament axis would be preserved in both situations. The inset of **l** depicts the vectors for a shifting transition of tropomyosin: an azimuthal and longitudinal shift of 12 Å and 35 Å, resulting in an overall shift of 37 Å.

Extended Data Table 1 | Data collection and refinement statistics

Data collection			
	Magnification	x122,270	
	Defocus range (μm)	0.8-2.6	
	Voltage (kV)	300	
	Microscope	Titan Krios (Cs corrected)	
	Camera	Falcon 2 (4k DED)	
	Frame recording time (s)	0.085-0.475	
	Number of frames	7	
	Electron dose (e$^-$Å^{-2})	15.4	
	Pixel size (Å)	1.12	
Particle statistics			
	Filaments*	7,854 (5,239)	
	Box size (px)	256	
	Boxing distance (px)	29	
	Rise (Å)	27.5	
	Azimuthal rotation (°)	166.4	
	Segments*	109,242 (74,228)	
	Asymmetric subunits*	~180,000 (~120,000)	
			
Model composition	**Chain ID**	**ABCDE**	**A**
	Non-hydrogen atoms	14,450	2,890
	Protein residues	1,835	367
	Ligand (ADP, Mg^{2+})	140	28
Refinement			
	Resolution used for refinement (Å)	95.2-3.7	
	Map sharpening *b* factor (Å^2)	-50	
	Average B-factor (Å^2)	55.40	53.40
	R factor†	0.277	
	Cross-Resolution function (FCR)‡	0.81	0.83
Rms deviations			
	Bonds (Å)	0.004	0.005
	Angles (°)	1.10	1.12
Validation			
	Clashscore, all atoms	19.99	18.32
	Poor rotamers (%)	1.23	1.29
Ramachandran plot			
	Favored (%)	87.20	87.40
	Outliers (%)	1.37	1.37

* Number in parentheses is count after sorting the dataset regarding micrograph quality

† R factor = $\Sigma |F_{obs}|-|F_{calc}| / \Sigma |F_{obs}|$, where F_{obs} describes the structure factors for all measured reflections and F_{calc} corresponds to the calculated structure factors from the final model

‡ FCR = $\Sigma(N_{shell}\ FSC_{shell}) / \Sigma N_{shell}$, with N_{shell} as number of structure factors in given shell and $FSC_{shell} = \Sigma(F_{calc}\ F_{EM}) / \sqrt{(\Sigma |F_{calc}|^2\ \Sigma |F_{EM}|^2)}$, where F_{EM} describes the structure factors for the final cryo-EM density map

Refinement statistics are shown after the last step of the refinement of the F-actin pentamer. In addition, relevant values of the central chain A are shown. Rms, root mean square.

CORRIGENDUM

doi:10.1038/nature14149

Corrigendum: Mutant IDH inhibits HNF-4α to block hepatocyte differentiation and promote biliary cancer

Supriya K. Saha, Christine A. Parachoniak, Krishna S. Ghanta, Julien Fitamant, Kenneth N. Ross, Mortada S. Najem, Sushma Gurumurthy, Esra A. Akbay, Daniela Sia, Helena Cornella, Oriana Miltiadous, Chad Walesky, Vikram Deshpande, Andrew X. Zhu, Aram F. Hezel, Katharine E. Yen, Kimberly S. Straley, Jeremy Travins, Janeta Popovici-Muller, Camelia Gliser, Cristina R. Ferrone, Udayan Apte, Josep M. Llovet, Kwok-Kin Wong, Sridhar Ramaswamy & Nabeel Bardeesy

Nature **513,** 110–114 (2014); doi:10.1038/nature13441

In this Article, the last line of the Acknowledgements should read: J.M.L. and D.S. are supported by the Asociación Española Contra el Cáncer (AECC); this has been corrected in the online versions of the paper.

CORRECTIONS & AMENDMENTS

ERRATUM

doi:10.1038/nature14150

Erratum: Deconstructing transcriptional heterogeneity in pluripotent stem cells

Roshan M. Kumar, Patrick Cahan, Alex K. Shalek, Rahul Satija, A. Jay DaleyKeyser, Hu Li, Jin Zhang, Keith Pardee, David Gennert, John J. Trombetta, Thomas C. Ferrante, Aviv Regev, George Q. Daley & James J. Collins

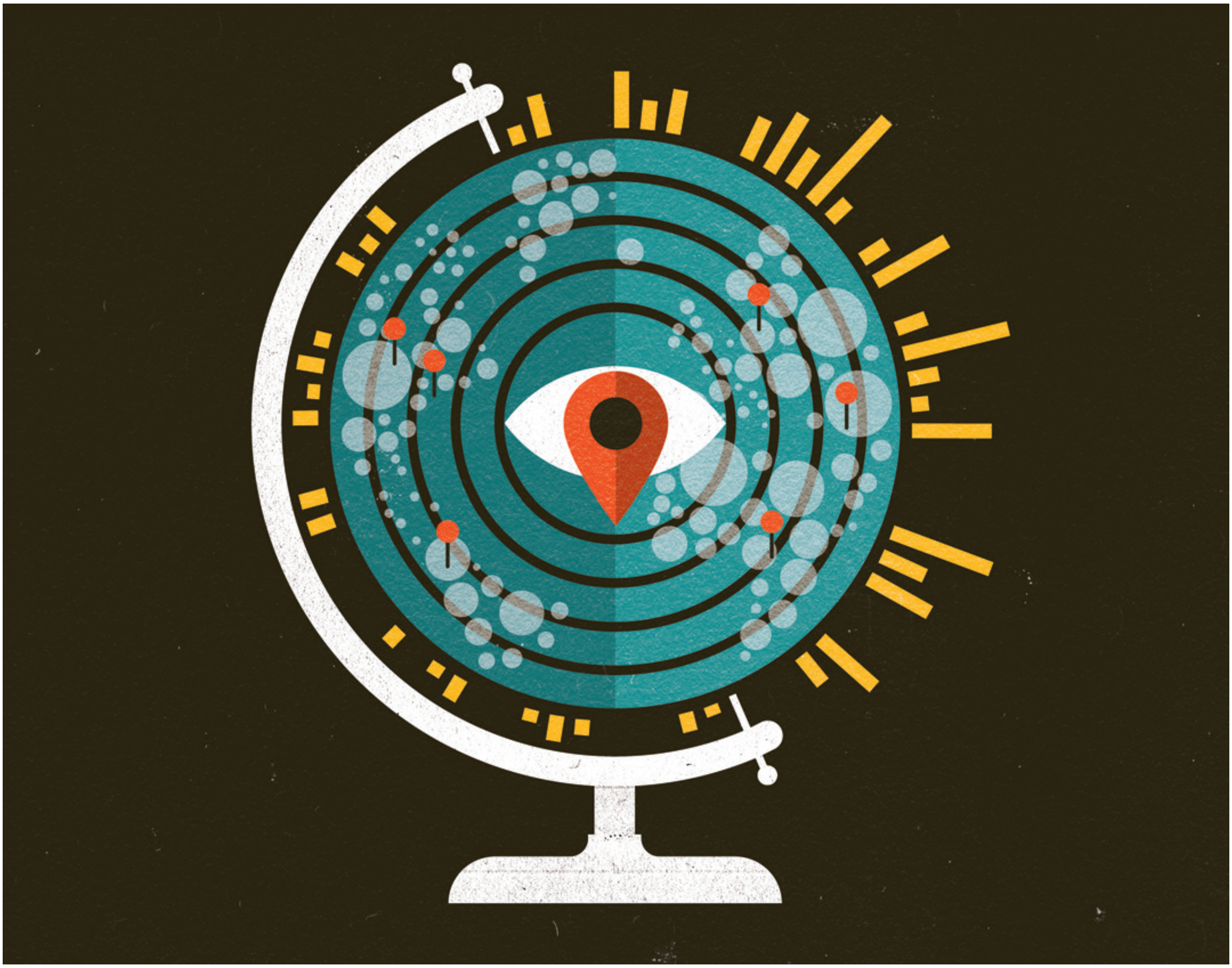
Nature **516,** 56–61 (2014); doi:10.1038/nature13920

In this Letter, owing to a production error, in Fig. 3e the sixth bar from the left was inadvertently labelled '*Dgcr8*$^{-/-}$ FF' instead of '*Dicer*$^{-/-}$ FF'. This has now been corrected in the online versions of the paper.

SCIENCE ON THE MAP

Easy-to-use mapping tools give researchers the power to create beautiful visualizations of geographic data.

ILLUSTRATION BY THE PROJECT TWINS



BY MARK ZASTROW

When linguist Lauren Gawne roams the valleys of Nepal documenting endangered Tibetan languages, she takes pains to distinguish each dialect's geographical origin. But when it came to producing maps of her results, for many years her cartographic methods were somewhat crude.

"My old maps were [made] using MS Paint on top of some copyrighted map that I really shouldn't have been using," she says. Her next solution wasn't much better: "My mum tracing a map off an atlas so that I had something a bit cleaner to work with." The one after that — "using Google Earth and dropping pins on it" — was generic, ugly and "looks horrible in a PowerPoint".

So in 2013, she jumped at the chance to join a workshop on mapping and visualization at the University of Melbourne in Australia, where she was working on her PhD. There she discovered the free, open-source program TileMill, created by the company Mapbox, which has offices in San Francisco, California, and in Washington DC. It lets users create maps from their data and pre-existing online cartographic databases.

➲ NATURE.COM
For more on scientific software, apps and online tools, visit:
nature.com/toolbox

TileMill is just one tool in the emerging field of customized mapping, where a bevy of open-source technologies and start-ups have given rise to an abundance of offerings for researchers and enthusiasts. These tools are more approachable for novices than the conventional geographic information systems (GISs) that geographers have long used for analysis of geospatial data sets. They allow non-specialists ▶

to easily visualize, manipulate and share their data in formats that are as slickly browsable as Google Maps but with greater power and flexibility. “TileMill allows you to be a complete control freak,” says Gawne, now at Nanyang Technological University in Singapore. From line styles to font spacing and kerning, “I can really manipulate all the variables quite easily.”

Until recently, Google, which is based in Mountain View, California, itself had staked the biggest claim in this space, providing various ways to access and decorate its maps through application programming interfaces (APIs). But as demand grew, the tech giant began limiting public access to its APIs in 2011 — and this allowed slightly more sophisticated open-source tools to flourish, says Oliver O’Brien, a geographer at University College London. Today, a fully fledged ecosystem of start-ups with open-source technology at their core offer platforms that many say have surpassed Google’s offerings.

“Google really nailed down having maps on the web,” says Javier de la Torre, a founder and current chief executive of one of Google’s emerging rivals, CartoDB of New York City. “What I think they didn’t see coming was that there was going to be this explosion of new map-makers.”

THE NEW MAPPING LANDSCAPE

In 2011, de la Torre was part of a team researching biodiversity informatics. The group was seeking an online platform to make a map of all known species on the planet. “There wasn’t technology for doing that,” he says — no tool could handle the amount of data, nor visualize how they changed over time.

The researchers decided to develop the tool themselves and created what became the open-source platform CartoDB. The company offers free and paid plans for hosting and visualizing data through its website. Unlike TileMill, which is primarily intended for drawing and designing static maps, CartoDB specializes in visualizing dynamic layers of data on top of basemaps. Users can import their geo-located data into CartoDB’s web-based interface and then filter or cluster data points, change the colour or size of symbols, and animate data changes over time. “CartoDB wants to be a place where your data lives,” says Steve Bennett, a research-oriented technologist at the University of Melbourne who takes workshops on mapping, including the one that Gawne attended.

Peter Desmet, who collaborates with a bird-tracking research team at the Research Institute for Nature and Forest in Brussels, was a colleague of de la Torre and became an early adopter of CartoDB. “I was never a desktop GIS person,” he says. But in CartoDB, “you can create and share a visualization in literally minutes”. Being able to simply send a link to the map online also makes it much faster to point out data-quality issues to colleagues, he says.

Another strength of CartoDB is its selection of global basemaps — ranging from familiar geopolitical and satellite-image formats to more stylish black-and-white and even pencil- and watercolour-themed renditions. Some are produced by TileMill’s maker Mapbox, which boasts a growing list of corporate and media clients — in many cases supplanting Google in a growing ‘battle of the basemaps’.

Mapbox first released TileMill in 2011. The team took a powerful but complex open-source cartographic renderer called Mapnik, built an easy-to-use interface around it and created a simple styling language, CartoCSS, to customize the maps’ appearance.

“TileMill was a game-changer, absolutely,” says Bennett. It allowed non-experts to produce professional-looking maps — either for publication as static figures or for use as basemaps in other visualization tools — without the need for more-complicated GIS programs.

The landscape continues to shift rapidly. In January, Google announced that it would shut down some premium and paid forms of Google Maps and focus on its basic Maps API. In response, CartoDB introduced tools to help users migrate their data to CartoDB, while still allowing them to integrate the Google Maps APIs. Mapbox, for its part, has shifted development from TileMill to its intended replacement, Mapbox Studio.

Cost of data storage is a potential stumbling block for scientists with large data sets — although CartoDB is open source, its convenience comes in large part from using it on the company’s hosted web service. The firm offers 75 megabytes of storage for free, but to store more than 1 gigabyte of data, the price rises quickly to hundreds of US dollars per month. CartoDB also charges to keep data and maps private on the site. “We’ve had real problems,” says Bennett. “If you’re a PhD student with no funding, it just doesn’t work.” Mapbox works with a similar pricing model for hosting maps on its servers, although TileMill itself is a free, downloadable program. However, CartoDB does work with academic users to try to find a solution, says de la Torre, and awards grants of up to US$3,500 to researchers studying the impacts of climate change, in recognition of the company’s environmental roots.

Power users can daisy-chain these tools together: for example, one could create a basemap in TileMill and data layers in CartoDB, then wrap them in an online interface using Leaflet, a mobile-friendly visualization package that runs in the program JavaScript and meshes with other JavaScript visualization packages such as D3. Duncan Smith, a geographer at University College London, has made one such combination: an online map of UK census data called LuminoCity that uses Leaflet to display the map data over basemaps produced in TileMill, and a variant of D3 called Dimple to show graphs of the data onscreen.

➲ MORE ONLINE

DUNCAN A. SMITH, CASA UCL

For more on scientific mapping, including a gallery and links to tools, see go.nature.com/u76knj

STORAGE HUBS

Researchers can also store their data sets in a CartoDB account, then access them (using the ubiquitous SQL database language) for other online applications, notes Desmet. For one project, he used D3 to build a map depicting radar observations of bird migration as wind-like flowing curves. The source code is stored in the repository GitHub, but the map pulls the scientific data from his CartoDB account.

Despite the visual sophistication of these tools, the level of computational analysis they provide is limited. But after using these programs to get to grips with the basic principles, researchers can progress to more-powerful GIS platforms. Many scientists — including those involved in public policy, such as urban planning and crisis mapping — use arcGIS, a suite of products maintained by Esri, based in Redlands, California. But there is also an open-source alternative: QGIS, a project of the Open Source Geospatial Foundation.

Researchers who already write code as part of their work can use programming languages such as Python and R, which already have capable mapping packages that users may not even be aware of, points out astronomer James Davenport of the University of Washington in Seattle. He says that astronomers often “end up bastardizing scientific visualization software to make maps”. He now uses the Python package matplotlib in tandem with the rest of his Python-based analysis to project his infrared observations onto maps of the sky.

Even researchers who would rather not touch a line of code can accomplish a lot with the help of CartoDB and TileMill. “You don’t have to be particularly technically competent,” says Gawne, who produced the Tibetan-language maps for her thesis in TileMill and now teaches mapping workshops herself. “You have to be not afraid to try it.” ■

Mark Zastrow *is a science writer in Seoul. He reported this article from Washington DC.*

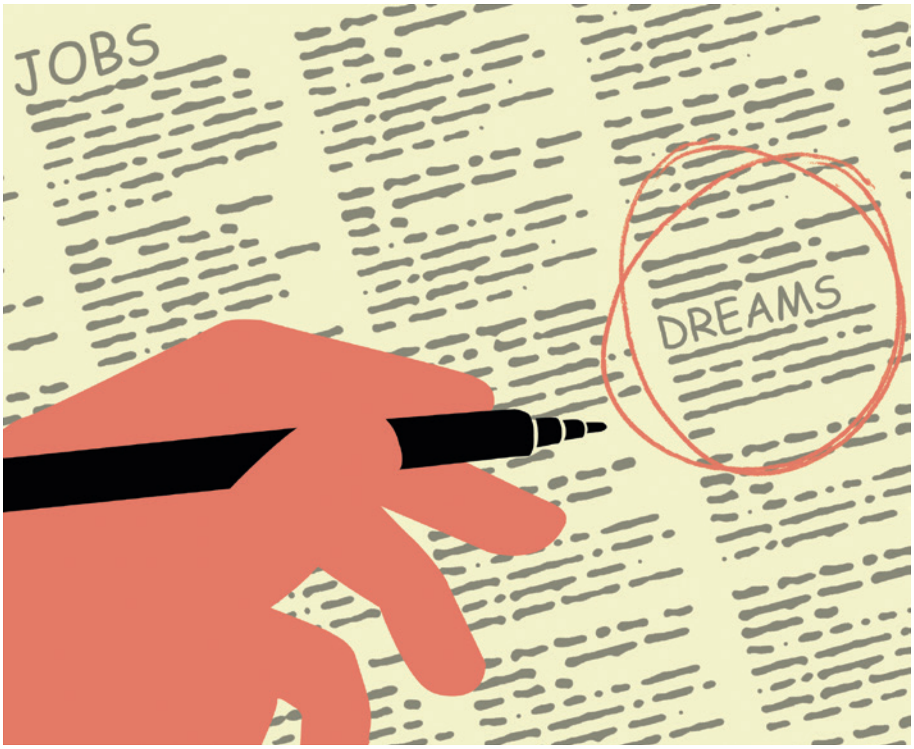
CLARIFICATION

The Toolbox story ‘Adventures with R’ (*Nature* **517,** 109–110; 2015) did not make clear what stopped Rabih Murr from practising R — he was preparing a paper for publication.

STARTING OUT Efforts to track PhD careers get under way slowly **p.122**

POSTDOCS Check out our new series on insights and options **go.nature.com/unxrhx**

OUTREACH How classroom work can yield insight in the lab **go.nature.com/8e8t64**



NEIL WEBB/GETTY

COLUMN

Wanted: information

Detailed career data will help people to plan for life after a PhD, say **Viviane Callier** and **Nathan L. Vanderford**.

Most students who enrol in US science and engineering PhD programmes hope to pursue an academic career. However, the gulf between the supply of newly minted PhDs and the availability of faculty positions widens each year. Some 36,000 people earned science and engineering PhDs in the United States in 2011, but US universities create only around 3,000 tenure-track positions annually. And with about 70% of those graduates taking a postdoctorate (M. Schillebeeckx, B. Maricque and C. Lewis *Nature Biotechnol.* **31,** 938–941; 2013), many trainees end up in a holding pattern, waiting for faculty jobs that are unlikely to materialize.

This is because the higher-education sector has not delivered an essential component of an efficient market: current and precise information about job prospects, including the specific attributes and training that have enabled PhD holders to find success in and outside academia, and the differences in those job markets for science, technology, engineering and mathematics (STEM) subfields.

Although some trainees may be told that faculty positions are a long shot, self-delusion is often part of the decision to pursue lengthy postdoctoral fellowships. After all, those who go to graduate school undeterred were often at the top of their undergraduate class, and believe they can 'beat the odds' of the job market. They have never experienced what it is like to be average — surrounded by equally bright peers.

Of course, the research enterprise colludes in this self-delusion, because it would not survive without the cheap labour supplied by graduate students and postdocs. In addition, trainees may have the intellectual ability to excel in academia, but many are left unprepared for the financial, psychological and personal costs of being on the academic job market for several years. Nor do many of them fully grasp the cost of waiting for a job that might never materialize. Students may remain woefully uninformed about alternatives, because there is even less information about the job market for non-faculty careers, or about the combination of skills and experience that are required to land a job outside academia.

Without better information about the specific set of qualifications, skills and experience required for finding a tenure-track or non-academic job in individual STEM fields, the job market is unlikely to be self-adjusting. Institutional demand for cheap research labour will continue to pull in hordes of graduate students and postdocs, even though the demand for tenure-track faculty members is vanishingly small.

Those pursuing a PhD need a more accurate picture of the academic and non-academic job markets, and they need it well before they graduate. This would provide a smooth transition rather than a rude awakening upon graduation. Federal agencies such as the US National Institutes of Health (NIH) and the US National Science Foundation (NSF) provide support for most US graduate students and postdocs, and they should also step up as information brokers. They should track career outcomes for academics and beyond, and gather data to compare people with and without PhDs in similar industries or capacities so that potential candidates can assess whether a PhD programme is likely to pay off. This information will not only benefit trainees, but also help federal agencies to manage the investment of government money in training STEM graduates. When graduates have limited career options that match their training, the return on investment is suboptimal.

Trainees need to know what employers are looking for. For faculty jobs, publications, gender and institutional prestige play a part, but so do complementarity with potential colleagues' research and an institution's orientation towards research, teaching and diversity. The criteria for hiring in non-academic careers are different, with an emphasis on transferable skills such as leadership, communication ▶

and teamwork. Non-academic employers also value work experience — sometimes more than academic credentials — and trainees who are interested in non-academic careers should gain this experience early on to avoid the common PhD catch-22 of being labelled simultaneously over- and under-qualified.

For the sake of future scientists, information about the current and projected state of the job market should be regularly collected, analysed and disseminated. Universities should curate data about former trainees. For academics, that information should include their grant, publication and teaching records as well as outreach and mentoring activities, and the criteria that academic hiring committees use to evaluate candidates. For non-academics, data should include the training, internships and work experience that led to employment.

The US National Postdoctoral Association in Washington DC is collecting data about postdocs' career paths as well as about institutional compensation, benefits and career services (see page 122). The NSF this winter launched an Early Career Doctorates survey that will gather in-depth information about postdocs and others who have earned their doctorates within the past ten years. The NIH's newly created Division of Biomedical Research Workforce in Washington DC may become the ideal organization for gathering and disseminating data about PhD graduates in the biomedical sciences.

These data will help funding agencies to craft policies that encourage institutions to give people with PhDs options for careers in non-faculty positions. One way to do this is to provide diverse education and training to PhD trainees so they can pursue careers in industry, consulting, entrepreneurship, science policy, writing and editing, administration or management. Federal funding agencies must find ways to ease pressure on trainees to work day and night for publications and grants, and instead foster ways to gain work experience and explore non-academic career paths while still in training.

These long-term solutions will not help current graduate students and postdocs, who must seek professional-development counselling, develop transferable skills and network within and outside academia. Ultimately, the careers of hundreds of thousands of future PhD holders depend on access to career information that will help to better match supply with demand. ■

Viviane Callier *is a research scholar at the Ronin Institute for Independent Scholarship in Montclair, New Jersey.* **Nathan L. Vanderford** *is an assistant professor and administrator at the University of Kentucky in Lexington.*

POSTGRADUATE CAREERS

The hunt for the elusive alumni

New efforts aim to track the careers of postgraduates.

BY PAUL SMAGLIK

CLAIRE WELSH/NATURE

Finding a job after finishing a PhD can be a tough slog. But a handful of US- and UK-based data-gathering initiatives could help science PhD holders to make more-informed decisions about their career options.

Previous efforts to track post-PhD career outcomes have focused mainly on one-off surveys, which usually offer little information beyond broad trends. Many also exclude postdoctoral researchers, who are difficult to follow if they work under a principal investigator's grant or are employed under a different job title.

In January, the US Council of Graduate Schools (CGS) in Washington DC published a report, *Understanding PhD Career Pathways for Program Improvement*, that calls for better measurement of postgraduate career outcomes. It estimates that about half of PhD graduates in science, technology, engineering and maths (STEM) find their first jobs outside academic institutions, but little other information is available. "We do not know the specifics of their careers: the nature and kinds of work produced and their long-term trajectories," the report says. It asks member institutions to use social media to locate and to conduct surveys of alumni.

Some US institutions, including Vanderbilt University in Nashville, Tennessee, have already initiated alumni-tracking schemes. Roger Chalkley, who heads Vanderbilt's biomedical research education and training programme, has started tracking the career paths of postdocs and graduate students using publication records and LinkedIn, among other social-media sites.

Vanderbilt's tracking results broadly mirror the trends reported by the CGS. About half of the university's STEM alumni stay in academic positions, although only about 20% land tenure-track jobs.

The other PhD graduates take on teaching, administration or laboratory-management positions, says Chalkley. "There's been an increasing flow of postdocs going to non-tenure research tracks."

Social-media sites are a key component of a global tracking venture launched by Vitae, a UK-based careers-support organization for researchers. Rather than seeking quantitative data about careers — the percentage of scientists who go into academic, industrial or government positions, for example — Vitae is gathering information about the career paths of specific individuals.

This approach is useful, says Vitae head Janet Metcalfe, because the biggest information gap is in what junior scientists do after their postdoctoral research — particularly if they do not land an academic position. "That's what people really want to know — if I don't make it on the academic route, what are my prospects?" she says.

The venture's first phase drew around 600 respondents: half were from the United Kingdom and the rest came from all over the world. Vitae will publish its results this month, although it has already posted several dozen stories of respondents' career paths online; these stories show that scientists can find professional fulfilment outside academic institutions, Metcalfe says. A second phase is under way: the results will be released this summer.

Longitudinal studies that follow large groups of scientists over time can present a clear picture of career progression, says Patrick Mulvey, a statistician at the American Institute of Physics in College Park, Maryland.

This winter, the US National Science Foundation launched just such a study, which will follow a cohort of scientists who earned their doctorates in the past decade. The study aims to bridge the gap between conventional surveys, which use rigorous sampling, and more-qualitative surveys, which emphasize individual stories.

When it delivers its results in 2016, the study should provide a cohesive picture of why each participant made their career choices and how they feel about them. But for now, the hunt for the elusive alumni continues. ■

SIMON CLASH: THE GALAXY'S GREATEST HERO

All in a day's work.

BY JAMES AQUILONE

A *brzzt-brzzt-brzzt* came stuttering through the air as we sped over the black sands of Desolation. Something hit the repulsor-cycle's rear fin, sending it into a vicious spin. Ja-bot was immediately thrown. I fought to regain control, and had nearly stopped the ever-widening gyre, but the cycle caught the edge of a dune and pitched me into the burning sand.

I was unhurt, beside what appeared to be a pointy shard of bone poking out just above my boot. I stood, tightened my laces, and watched as a sand skiff swooped down from the supernova-bright sky. Harvesters! The organ-jackers must have been desperate if they were operating on this miserable rock.

I gripped my sonic wand, and when the six Harvesters leaped out of their skiff, I gave it a sharp flick — which should have produced a long, pulsating laser whip. Instead, it coughed up a limp noodle of a stream that died at the Harvesters' feet. My trusty sidekick had only one responsibility before we set out to save Princess Velouria from the Dominion: to charge the damn wand!

"Ja-bot, you waste of metal!"

"My sincerest apologies, Master Clash."

Oh well, a hero of my calibre doesn't need the most powerful and versatile weapon in the Galaxy. My fists have been declared weapons of astronomical destruction in seven systems. I cracked Harvester jaws and noses with the ferocity of a three-headed Gandavian megapig, and the tiny, enshrouded creatures dropped at my feet. Except one. How the runt knocked me out with an electro-stunner, I'll never know. Ja-bot must have been blocking my view.

I awoke minutes later staring up at Desolation's triplet suns. I tried to move, but it was futile. I was pinned, naked, to a biomagnetic surgical table on the skiff's deck. Ja-bot lay beside me. The Harvesters had already disconnected his head. Unfortunately it was still operable.

"It's hopeless, Master Clash," Ja-bot said. "We're doomed!"

"A hero always finds a way, Ja-bot. That's what makes him a hero."

"Of course, Master Clash. Forgive my lack of faith."

A Harvester stood over me with a plasma scalpel. In his pipsqueak voice, he said: "Your organs should fetch a good price on the Ondorean black market. It's rare for a human to be found on Desolation."

The Harvester severed my left arm at the shoulder. Thanks to the surgical table, there was no pain, no blood.

"Sacrifice, Ja-bot. It's all part of the hero's struggle." I glared at the butcher. "Is that all you got, Harvester scum!"

It wasn't. The Harvester raised the plasma scalpel again.

ILLUSTRATION BY JACEY

The hero faces many challenges before his eventual triumph. What was one more?

The scalpel slid about three centimetres into my shoulder before bright bursts of purple erupted before my eyes.

When my vision returned, I was looking up at a masked figure who was nearly twice the size of a Harvester and wore the black and silver uniform of the Dominion. The Harvesters lay on the ground, both cut neatly in half.

"Dominion scum!" I said. "My day is finally looking up."

The figure slowly shook its head and then removed its mask.

"Princess Velouria!"

"Clash, you've seen better days."

"This? You should have seen me after I saved Princess Dreakia. You don't want to know what they cut off me then."

Suddenly the other four Harvesters appeared from below deck and rushed towards the princess.

"Free me, princess," I said, "so that I can rescue you."

But she must not have heard me in her terror and panic. Instead, she drew a sonic wand from her back pocket. With a flick of her wrist, a thick rope of pulsating purple light lashed out and sliced through three of the Harvesters. They fell to the ground in six pieces. The fourth Harvester leaped over the laser whip, somersaulted in the air, and with an electro-stunner in his hand, bore down on the princess.

With amazing luck, she slid to the side at the precise moment, snatched the Harvester's foot, slammed him to the ground, and with a shake of her wand fired a laser blast into his head.

The princess wiped a piece of the Harvester's brain matter off her cheek. "What was that about a rescue, Clash?"

I didn't dignify her with an answer. The princess was well known for her inappropriate sense of humour.

"We have an hour to get off Desolation," she said. "I've sent a self-tunnelling antimatter bomb to the centre of the planet."

"Why the hell would you do that?"

"After I rescued myself the first time, I discovered that Desolation is a living cosmic being. The Dominion plan on using it to destroy the Confederation."

"Oh, then you'll be needing a ride off this doomed body?"

"I was on my way to the space elevator, but I've probably wasted too much time here."

"Simon Clash is my name, rescuing princesses is my game. My ship is just two klicks to the south. I'll have you safely back in the Confederation's arms in no time."

"Speaking of arms. Don't forget yours."

Princess Velouria had retired to my ship's sleeping quarters by the time Desolation went critical. The poor thing must have been exhausted after her harrowing abduction. She had said she was on Desolation as a spy. I chalked that nonsense up to delirium.

"Another princess saved, another evil empire defeated," I said and flexed my left hand. Ja-bot had restored me in the medic bay. At least he was good for something.

"You truly are the Galaxy's greatest hero, Master Clash."

"No, no, Ja-bot, true heroes don't take credit for their heroic deeds. We let our actions speak for themselves." ■

James Aquilone *is an editor and writer. Visit his website at jamesaquilone.com.*

➲ NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm